From (10) and (14) we conclude that the joint distribution density of the real and imaginary parts of the roots of (9) is given by

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{2n} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{j=1}^{n} z_j \sum_{j=1}^{n} \bar{z}_j + \cdots + z_1\bar{z}_1 \cdots z_n\bar{z}_n\right\}\right] \sum_{p=1}^{n} \sum_{q=p+1}^{n} | z_p - z_q |^2. \tag{15}$$

A NOTE ON THE PROBABILITY OF ARBITRARY EVENTS

By Hilda Geiringer[1]

Bryn Mawr College

In a recently published paper [1] on arbitrary events the author studies the probability of the occurrence of at least m among n events. Denoting by $p_m(\gamma_1, \gamma_2, \cdots \gamma_r)$ the probability that at least m among the r events, $E_{\gamma_1}, \cdots E_{\gamma_r}$ occur, and by $p_{[\alpha_1,\alpha_2,\ldots,\alpha_r]}$ the probability of the non occurrence of the events numbered $\alpha_1, \alpha_2, \cdots \alpha_r$ and of the occurrence of the $n - r$ others, he proves

$$-p_1(\alpha_{r+1}, \cdots \alpha_n) + \sum_{\gamma_1} p_1(\gamma_1, \alpha_{r+1}, \cdots \alpha_n) - \sum_{\gamma_1}\sum_{\gamma_2} p_1(\gamma_1, \gamma_2, \alpha_{r+1}, \cdots \alpha_n) + \cdots + (-1)^r \sum p_1(1, \cdots n) = p_{[\alpha_1, \cdots \alpha_r]}. \tag{I}$$

(Theorem VI, page 336). From (I) he deduces that *a necessary and sufficient condition* for the existence of a system of events $E_1, \cdots E_n$ associated with given values $t_1(\alpha_1, \cdots \alpha_k)$ is that the expressions on the left side of (I) computed from these t's are $\geqq 0$ for all possible combinations of the α's (Theorem VII). He also points out that it was not possible to find similar (necessary and sufficient) conditions for $m \neq 1$. I wish to show in this note the relation between these theorems and some well known basic facts of the theory of arbitrarily linked events and to add some remarks.

1. Given n chance variables x_i $(i = 1, \cdots n)$ denote by $x_i = 1$ the "occurrence of E_i", by $x_i = 0$ its non occurrence and by $v(x_1, x_2, \cdots x_n)$ the probability of "the result $(x_1, x_2, \cdots x_n)$" i.e., the probability that the first variable equals x_1 the second x_2, $\cdots$ the last x_n; e.g. $v(1, 1, 1, 0, \cdots 0) = v_{[45\ldots n]}$ is the probability that only the three first events occur. Hence the v's are 2^n probabilities, arbitrary except for the *condition to have the sum* 1.

Instead of these v's we often introduce another set of $2^n - 1$ probabilities, namely p_i the probability of the occurrence of E_i $(i = 1, \cdots n)$; p_{ij} that of the joint occurrence of E_i and E_j $(i, j = 1, \cdots n)$; $\cdots$ $p_{12\ldots n}$ the probability that *all* the events occur.

[1] Research under a grant-in-aid of the American Philosophical Society.

It may be noted that instead of the p_i, p_{ij}, $\cdots p_{12\ldots n}$ we could quite as well use a system of q_i, q_{ij}, $\cdots q_{12\ldots n}$ where q_i is the probability of the non-occurrence of E_i (or of the occurrence of $E'_i = E - E_i$, q_{ij} that of the joint non-occurrence of E_i and E_j (of the occurrence of $E'_iE'_j$) and $q_{12\ldots n}$ the probability of $E'_1E'_2 \cdots E'_n$.

The use of the p's (or q's) instead of the "elementary probabilities" v is justified by the fact that the p's are $(2^n - 1)$ *independent* linear combinations of the v's and that therefore *the v's and the p's* (or the v's and the q's) *determine each other uniquely*. There exist in fact the following well known relations, (1) and (2). The first set (1) gives just the definition of the $2^n - 1$ probabilities p, in terms of the v's, and the second set expresses the v's by the p's as the result of the solution of the $2^n - 1$ independent linear equations (1). Thus we have, beginning with $p_{12\ldots n}$:

$$
\begin{aligned}
p_{12\cdots n} &= v(1, 1, \cdots 1),\\
p_{12\cdots n-1} &= \sum_{x_n} v(1, 1, \cdots 1, x_n),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
p_{12} &= \sum_{x_3} \cdots \sum_{x_n} v(1, 1, x_3, x_4, \cdots x_n),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
p_n &= \sum_{x_1} \cdots \sum_{x_{n-1}} v(x_1, x_2, \cdots x_{n-1}, 1),
\end{aligned}
\tag{1}
$$

and solving successively:

$$
\begin{aligned}
v(1, 1, \cdots 1) &= p_{12\cdots n}\\
v(1, 1, \cdots 1, 0) &= p_{12\cdots n-1} - p_{12\cdots n},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
v(1, 1, 0, \cdots 0) &= p_{12} - \sum_{\gamma_1} p_{12\gamma_1} + \sum_{\gamma_1}\sum_{\gamma_2} p_{12\gamma_1\gamma_2} - \cdots \pm p_{12\cdots n},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
v(0, 0, \cdots 0, 1) &= p_n - \sum_{\gamma_1} p_{\gamma_1 n} + \sum_{\gamma_1}\sum_{\gamma_2} p_{\gamma_1\gamma_2 n} - \cdots\\
&\qquad \pm \sum_{\gamma_1} \cdots \sum_{\gamma_{n-1}} p_{\gamma_1\gamma_2\cdots\gamma_{n-1}n} \mp p_{12\cdots n}.
\end{aligned}
\tag{2}
$$

The successive solution of the system (1) with respect to the "unknowns" v is possible because each new equation in (1) contains *exactly one* new unknown v; e.g. in the equation defining p_{12} the only "unknown" is $v(1, 1, 0, 0, \cdots 0)$ all the v's with more than two "1"s having already been computed from the foregoing equations.

If we choose to use the system of the q's we have in the same way:

$$
\begin{aligned}
q_{12\cdots n} &= v(0, 0, \cdots 0),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
q_n &= \sum_{x_1} \cdots \sum_{x_{n-1}} v(x_1, x_2, \cdots x_{n-1}, 0),
\end{aligned}
\tag{1'}
$$

and the inverse system

$$(2') \quad \begin{aligned} v(0, 0, \cdots 0) &= q_{12\cdots n} \\ v(0, 0, \cdots 0, 1) &= q_{12\cdots n-1} - q_{12\cdots n} \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ v(1, 1, \cdots 1, 0) &= q_n - \sum_{\gamma_1} q_{\gamma_1 n} + \sum_{\gamma_1}\sum_{\gamma_2} q_{\gamma_1\gamma_2 n} - \cdots \mp q_{12\cdots n}. \end{aligned}$$

Coming back to Chung's theorem we see that the probability $p_1(\alpha_1, \cdots \alpha_r)$ that at least one event among $E_{\alpha_1}, \cdots E_{\alpha_r}$ occurs is evidently:

$$(3) \qquad p_1(\alpha_1, \cdots \alpha_r) = 1 - q_{\alpha_1,\dots\alpha_r}.$$

If we introduce this value in (I) all the "1"s introduced by (3) cancel and *we get our system* (2'). (Of course we could in the same way deduce from (2) a system of equations for $q_1(\alpha_1, \cdots \alpha_r) = 1 - p_{\alpha_1,\dots\alpha_r}$ where $q_1(\alpha_1, \cdots \alpha_r$ is the probability that at most $r - 1$ events among the r given ones occur.)

As the v-values on the left side of (2') are $2^n - 1$ independent probabilities only subject to the restriction that they have the sum $\leqq 1$, we see that the expressions on the right side of (2') *must have the same properties*; and these properties are also *sufficient* for a system of q's, (or for a system of $p_1(\alpha_1, \cdots \alpha_r)$; indeed *if* they are fulfilled, these $2^n - 1$ expressions define by means of (2') a system of elementary probabilities $v(x_1, \cdots x_n)$. Hence the theorems VI and VII quoted at the beginning of this note are rather close consequences of the basic relations of the theory of arbitrary events.

2. Remark 1. We may add one more equation to equations (1), namely

$$1 = \sum_{x_1} \cdots \sum_{x_n} v(x_1, \cdots x_n),$$

thus introducing $v(0, 0, \cdots 0)$. Then in system (2) the corresponding new equation will be:

$$v(0, 0, \cdots 0) = 1 - \sum_{\gamma_1} p_{\gamma_1} + \sum_{\gamma_1}\sum_{\gamma_2} p_{\gamma_1\gamma_2} - \cdots \pm p_{12\cdots n}$$

(and analogously for the q's). In this way we get two systems $(\bar{1})$ and $(\bar{2})$ each consisting of 2^n equations and in $(\bar{2})$ the sum of the expressions on the right side is now *identically equal* to one. Hence necessary and sufficient conditions will now be that *all these* 2^n *expressions must be non-negative.*

Remark 2. It is convenient to interpret or prove results of the kind considered here in terms of elementary measure theory: p_1 is the measure of a set E_1; p_2 of E_2; p_{12} that of the intersection E_1E_2 etc., and analogously for the v's: e.g. $v(1, 1, 0, \cdots 0) = m(E_1E_2E_3' \cdots E_n')$. Consider now the equations (2). The first is an identity. In the second $p_{12\dots n-1}$ measures the product of $E_1E_2 \cdots E_{n-1}$, whereas $p_{12\cdots n-1} - p_{12\dots n}$ is the measure of *that* part of this product which does not belong to E_n, and it therefore equals $m(E_1E_2 \cdots E_{n-1}E_n') = v(1, 1 \cdots 1, 0)$. In the last equation (2) $\sum_{\gamma_1} \cdots \sum_{\gamma_{n-2}} p_{\gamma_1\dots\gamma_{n-2}n}$ is the measure of that part of E_n which belongs to at least $(n - 2)$ other sets (besides E_n); whereas this same value minus $p_{12\cdots n}$ is the measure of the part of E_n which belongs *exactly* to $(n - 2)$ other sets; now subtracting this expression from $\sum_{\gamma_1} \cdots \sum_{\gamma_{n-3}} p_{\gamma_1\cdots\gamma_{n-3}}$ we get the measure of the part of E_n which belongs *exactly* to $(n - 3)$ other sets and finally $p_n - \cdots \mp p_{12\cdots n}$ is the measure of that part of E_n which belongs to *no* other set besides, i.e. $m(E_1'E_2' \cdots E_{n-1}'E_n) = v(0, 0, \cdots 0, 1)$. This kind of proof does not require the solution of (1).

REMARK 3. According to (1) the p_i, p_{ij}, $\cdots p_{12\cdots n}$ are the ordinary *moments* of order $1, 2 \cdots n$ of $v(x_1, x_2, \cdots x_n)$. There are of course many more than $2^n - 1$ moments of this n-variate distribution but only $2^n - 1$ of them are different from each other because $1^r = 1$.

3. Denote by $p_n(x)$, $(x = 0, 1, \cdots n)$ the *probability of getting exactly x successes in n trials.* (See e.g. [2], [3].) For the simplest case of arbitrary events, the Bernoulli problem, $p_n(x) = \binom{n}{x} p^x(1 - p)^{n-x}$. Then the probabiliy of *at least* x successes (of a number of successes $\geqq x$) is

$$V_n(x) = p_n(x) + p_n(x + 1) + \cdots p_n(n), \tag{4}$$

or $p_x(1, 2, \cdots n)$ in Chung's notation. The $p_n(x)$ are by their *definition* $(n + 1)$ *arbitrary positive* numbers with *sum equal to one.* These are the only *necessary* and *sufficient* restrictions for $p_n(x)$. $V_n(x)$ the "cumulative" distribution of $p_n(x)$ which is defined for x between $(-\infty$ and $+\infty)$ is a monotone non-increasing *step function* with its $(n + 1)$ steps at $x = 0, 1, 2, \cdots n$ equal to the $p_n(x)$.

Consider next $p_x(\alpha_1, \alpha_2, \cdots \alpha_r)$ where $r < n$; these are cumulative distributions each corresponding to one of the $\binom{n}{r}$ probabilities $p_r(x)$ where $p_r(x)$ is the probability of exactly x successes in a group of r trials.[2] For each group $(\alpha_1, \cdots \alpha_r)$ the corresponding $p_r(x)$, $(x = 0, 1, \cdots r)$ are positive and with sum equal to one. Hence if we always omit $p_r(0)$ because of $\sum_x^{0\cdots r} p_r(x) = 1$, all the different $p_1(x), p_2(x), \cdots p_n(x)$ together define

$$1\cdot n + n(n - 1) + \binom{n}{2}(n - 2) + \cdots + n = n2^{n-1}$$

values. As $n2^{n-1} > 2^n$ for $n > 2$ we realise that between these $n2^{n-1}$ probabilities there must exist a set of $n2^{n-1} - (2^n - 1)$ *identical relations*; and the same is true for the corresponding cumulative distributions $V_r(x)$ or $p_x(\alpha_1, \cdots \alpha_r)$. Thus it seems reasonable that it may be hard to use these $p_x(\alpha_1, \cdots \alpha_r)$ in the characterization of a problem of arbitrarily linked events if $x > 1$. On the other hand we have seen in **1** that for $x = 1$ they reduce to the

[2] One may write here $p_r(x)$ instead of $p_{(\alpha_1, \alpha_2, \cdots, \alpha_r)}(x)$.

$2^n - 1$ probabilities q_i, q_{ij}, $\cdots$ $q_{12\ldots n}$ which of course define the system of events unequivocally.

4. Introduce in the usual way the *sums* of the p_i, p_{ij}, etc.

$$(5) \qquad S_1 = \sum_i p_i, \qquad S_2 = \sum_{i,j} p_{ij}, \cdots S_n = p_{12\ldots n}, \quad \text{and} \quad S_0 = 1.$$

Now add in system (1) first the n equations which define p_1, p_2, $\cdots$ p_n, then the $\binom{n}{2}$ equations for the p_{ij}, etc. Observing that $p_n(x)$ is the sum of all *these* elementary probabilities $v(x_1\, x_2 \cdots x_n)$ with exactly x "1"s and $(n - x)$ "0"s we get as the result of these n additions the well known formulae:

$$(6) \qquad S_\gamma = \sum_{x=\gamma}^{n} \binom{x}{\gamma} p_n(x), \qquad (\gamma = 0, 1, \cdots n).$$

Here $\gamma = 0$ gives $S_0 = 1 = \sum_0^n p_n(x)$. We may solve successively these $(n + 1)$ linear equations with respect to $p_n(n)$, $p_n(n - 1)$, $\cdots$ $p_n(0)$, each linear equation containing only one new unknown, and find:

$$(7) \qquad p_n(x) = \sum_{\gamma=x}^{n} (-1)^{\gamma+x} \binom{\gamma}{x} S_\gamma, \qquad (x = 0, 1, \cdots n).$$

(These formulae could also have been derived from (2) by collecting groups of equations such that all the corresponding $v(x_1, \cdots x_n)$ contain the same number of "1"s.) (In the measure interpretation $p_n(x)$ is the measure of that part which belongs exactly to x of the original sets and S_γ measures the set which belongs to at least γ of these sets.) We also find by "cumulating" equations (9)

$$(8) \qquad S_\gamma = \sum_{x=\gamma}^{n} \binom{x-1}{\gamma-1} V_n(x), \qquad (\gamma = 1, 2, \cdots n),$$

and the inverse system

$$(9) \qquad V_n(x) = \sum_{\gamma=x}^{n} (-1)^{\gamma+x} \binom{\gamma-1}{x-1} S_\gamma, \qquad (x = 1, 2, \cdots n).$$

(6) and (8) are of the same type as (1), and (7) and (9) of the same as (2). We also may deduce analogous formulae by interchanging the roles of 0 and 1 and introducing a system of T_1, T_2, $\cdots$ T_n which depends on the q's in the same way as the S_1, S_2, $\cdots$ S_n defined in (5) depend on the p's.

We have seen that the $p_n(x)$ are $(n + 1)$ arbitrary non-negative numbers subject to only the condition of having their sum equal to one. But the S_γ $(\gamma = 0, \cdots n)$ are not arbitrary as we see from (7). The $(n + 1)$ expressions on the right side of (7) *must each be non-negative if* they are to define the probabilities $p_n(x)$ (their sum is identically equal to one). *Then and only then* they define a system of arbitrarily liked events E_1, $\cdots$ E_n.

The $p_n(x)$, $(x = 0, 1, \cdots n)$ are of course not equivalent to the complete system of $2^n - 1$ values $v(x_1, x_2, \cdots x_n)$ and the same remark holds for the $S_0, \cdots S_n$ and the system of $p_i, p_{ij}, \cdots p_{12\ldots n}$. But often we are particularly interested in problems dealing only with the $p_n(x)$ (and S_γ). (For instance the author has studied [2] the asymptotic behavior of $p_n(x)$ as n tends in different ways towards infinity.) The simplest way to indicate a particular p-system corresponding to given S_γ is of course to assume all the p_i equal to each other; all the p_{ij} equal to each other etc. and to put therefore:

$$p_1 = p_2 \cdots = p_n = \frac{1}{n} S_1,$$

$$p_{12} = \cdots = p_{n-1,n} = \left[1 \Big/ \binom{n}{2}\right] S_2, \cdots p_{12\ldots n} = S_n.$$

In the corresponding v-system all these v's which show the same number of "1"s equal each other.

We see from (6) that the S_γ (multiplied by $\gamma!$) are the *factorial moments* of order 0, 1, $\cdots$ n *of the distribution* $p_n(x)$. Therefore by (7) we get the $p_n(x)$ *in terms of their factorial moments up to order n*. We may therefore also say: *Necessary and sufficient conditions that a system of numbers* $N_0 = 1, N_1, \cdots N_n$ *be the factorial moment of an arithmetical distribution with at most* $(n + 1)$ *steps at* $x = 0, 1, \cdots n$ *are the inequalities*:

$$\sum_{\gamma}^{x \cdots n} \frac{(-1)^{\gamma+x}}{x!(\gamma - x)!} N_\gamma \geqq 0, \qquad (x = 0, 1, \cdots n). \tag{10}$$

Note that here there is no more allusion to a set of arbitrary events; (10) are the necessary and sufficient conditions for a set of $(n + 1)$ numbers to be the $(n + 1)$ (factorial) moments of an arbitrary arithmetic distribution with its abscissae given. The linear inequalities (10) differ very much from the basic inequalities in the classical problem of moments; because in our problem the abscissae of the steps are given in advance.

5. In some problems (e.g., some questions connected with the law of large numbers, with correlation theory, with analysis of variance) we are only concerned with the *first and second* moment of a distribution. Thus we are lead to the following question: Given $r + 1$ numbers $N_0, N_1, \cdots N_r$, $(r \leqq n)$ indicate a set of necessary and sufficient conditions such that these numbers are the moments of an arithmetic distribution with at most $(n + 1)$ steps, at $0, 1, 2, \cdots n$.[3] Some sort of an answer which may work well in particular cases, can immediately be deduced from (10). "$r + 1$ numbers $N_0, N_1, \cdots N_r$ will be the factorial moments of an arithmetic distribution with, at most, $(n + 1)$ steps at $0, 1, 2, \cdots n$ if and only if it is possible to indicate s

[3] This problem and the method of its solution has much in common with a problem studied in R. von Mises' paper [4].

numbers $N_{r+1}, \cdots N_{r+s}$, $(0 \leqq s \leqq n - r)$, such that for the $r + s + 1$ numbers $N_0, N_1, \cdots N_r, \cdots N_{r+s}$ the $r + s + 1$ inequalities

$$\sum_{\gamma=x}^{r+s} \frac{(-1)^{\gamma+x}}{x!\,(\gamma - x)!} N_\gamma \geqq 0 \qquad (x = 0, 1, \cdots r + s)$$

be satisfied."

The proof of this statement is self evident but the statement itself cannot be considered satisfactory. We get a general solution in the following way.

Let $f_1(t), \cdots f_r(t)$ be r functions of the chance variable t, $v(t)$ an arithmetic probability with n given attributes $t_1, t_2, \cdots t_n$ and

$$(11) \qquad E[f_\rho(t)] = \sum_1^n f_\rho(t_\gamma)v(t_\gamma) \equiv \sum_{\gamma=1}^n a_{\gamma\rho} v_\gamma = S_\rho, \qquad (\rho = 1, 2, \cdots r),$$

the expectations of $f_\rho(t)$ with respect to $v(t)$. We wish to indicate necessary and sufficient conditions for the r numbers S_ρ. For $f_\rho(t) = t^\rho$ we have the problem stated above where the first r moments are given.

Call (S) the r-dimensional curve $x_\rho = f_\rho(t)$ and $P_1, P_2, \cdots P_n$ the points on (S) with coordinates $f_\rho(t_\gamma) = a_{\gamma\rho}$, $(\rho = 1, \cdots r; \gamma = 1, \cdots n)$, S the given point with coordinates S_ρ. In this case, *the point S must be contained in the smallest convex body (B) determined by the n points $P_1, \cdots P_n$. This condition is necessary and sufficient.* Because, if we interpret the v_γ which are $\geqq 0$ as masses of the points P_γ, with sum equal to one, then S is the center of gravity of these masses and it is well known that the above mentioned condition for S has to be fulfilled. But this condition is also sufficient, because if S is contained in (B) there exists always a simplex of at most r dimensions, consisting of at most $(r + 1)$ of the given points such that S is the center of gravity of appropriate masses in these points.

If we want to indicate explicitly the inequalities for the S_ρ we must know the boundary of (B). This is determined by its *planes of support* ("Stützebenen," Minkowski) sometimes called *tack planes*. A tack plane is a plane which *does not separate* any two points of the given point set and contains at least one point of this set. A plane is said to *separate* two points if, when the coordinates of the points are written in the equation of the plane two values with opposite signs result. These definitions enable us to find those points P_γ which lie on the boundary of (B) and to determine this boundary. (E.g. for $r = 3$ we have to find such triples of i, k, l, that the determinant which represents the equation of the plane through these three points has the same sign for all possible other points P_λ. If the S_ρ are the first three *moments* with respect to the origin, these determinants become Vandermond determinants and we find easily that the boundary planes are each passing through two neighboring points P_γ, $P_{\gamma+1}$ and one of the endpoints P_1 or P_n. If $\rho = 2$, and the first two moments are given, the boundary of (B) consists of the polygon $P_1P_2 \cdots P_nP_1$). Then we find without difficulty the conditions to be satisfied by S in the form of *linear inequalities* between the given $S_1, S_2, \cdots S_r$.

We get the continuous case as a limit of the discontinuous case as $t_\gamma \rightarrow t_{\gamma+1}$

and the points $P(t)$ take up the whole curve (C), e.g. between $t = 0$ and ∞. Then the relations between the given S_ρ become *non-linear* inequalities, well known for the problem of moments.

REFERENCES

[1] KAI LAI CHUNG, "On the probability of the occurrence of at least m events among n arbitrary events," *Annals of Math. Stat.*, Vol. 12 (1941).
[2] HILDA GEIRINGER, "Sur les variables aléatoires arbitrairement liees," *Revue Interbalcanique*, 1938.
[3] HILDA GEIRINGER, "Bemerkung zur Wahrscheinlichkeit nicht unabhängiger Ereignisse," *Revue Interbalcanique*, 1939.
[4] R. v. MISES, "The limits of a distribution function if two expected values are given," *Annals of Math. Stat.*, Vol. 10, 1939.

AN INEQUALITY FOR MILL'S RATIO

BY Z. W. BIRNBAUM
University of Washington

Mr. R. D. Gordon[1] recently proved the inequalities

$$\frac{x}{x^2+1}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \le \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{1}{2}t^2}\,dt \le \frac{1}{x}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, \quad \text{for } x > 0.$$

In the present note we show that the lower inequality can be replaced by the better estimate

$$\frac{\sqrt{4+x^2}-x}{2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \le \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{1}{2}t^2}\,dt.$$

PROOF: According to a well-known theorem of Jensen[2], for $f(t)$ convex and $g(t) \geqq 0$ in the interval (a, b), the following inequality holds

$$f\left[\int_a^b tg(t)\,dt\Big/\int_a^b g(t)\,dt\right] \le \int_a^b f(t)\,g(t)\,dt\Big/\int_a^b g(t)\,dt.$$

For $a = x$, $b = \infty$, $f(t) = 1/t$, $g(t) = te^{-t^2/2}$, this inequality gives

$$\int_x^\infty te^{-\frac{1}{2}t^2}\,dt\Big/\int_x^\infty t^2e^{-\frac{1}{2}t^2}\,dt \le \int_x^\infty e^{-\frac{1}{2}t^2}\,dt\Big/\int_x^\infty te^{-\frac{1}{2}t^2}\,dt.$$

Since

$$\int_x^\infty te^{-\frac{1}{2}t^2}\,dt = e^{-\frac{1}{2}x^2} \qquad \text{and} \qquad \int_x^\infty t^2e^{-\frac{1}{2}t^2}\,dt = xe^{-\frac{1}{2}x^2} + \int_x^\infty e^{-\frac{1}{2}t^2}\,dt,$$

[1] R. D. Gordon, "Values of Mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 364–366.

[2] See for example: G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge, 1934, p. 150–151.