This note warns Computer Scientists and Engineers about hazards intrinsic in attempts to assess by probabilistic methods the seriousness of a design error after it has been discovered.

Whenever an engineer's mistake is discovered too late, engineering managers will require assessments of the mistake's impact before they can choose a course of remedial action. Often decisions are based upon probabilistic estimates of failure rates; lower estimates incur less desperate remedial activities. Unfortunately the lowest estimates tend also to be the least reliable, and not merely because they may be self-serving. Circumstances can confound the predictive power of low probabilities, and then decisions based upon them will fare worse than decisions reached after a frank acknowledgement of ignorance has led to the investigation of possibilities that would otherwise have been dismissed prematurely as altogether too improbable.

The kind of probabilistic reasoning customarily employed to predict failure rates due to random defects in telephone equipment and semiconductor fabrication cannot be employed confidently to predict failure rates due to a fault in the design of an arithmetic or logical device or software. Confidence is undermined by several considerations, foremost among them being the entirely non-random nature of that fault after it has been found. Usually, to proceed with probabilistic reasoning, we must assume that inputs are random, that operations occur with frequencies revealed by random sampling, that earlier malfunctions propagate their consequences into later operations at rates determinable by statistical methods, and so on. Were every such assumption a reasonably good approximation in its own right, yet their concatenation into a chain of inference would still be unreasonable, the more so if it predicted an extremely low failure rate. Predictions like that are often falsified by rare events that have been disregarded completely, and by correlations that have been neglected.

Unanticipated Rare Events

Calamities can be precipitated by unfortunate coincidences of design flaws with rare events. Here are three recent examples:

• Keeping a *Patriot* anti-missile battery in continuous action for more than a few hours during the Gulf War of 1991 let tiny timing discrepancies accumulate enough to degrade fatally the battery's interception capability. Even when the battery did intercept an incoming *SCUD* missile, its empty fuel tank was destroyed but its explosive warhead continued to its target.

• Rare freezing weather in Florida embrittled a synthetic rubber sealing ring on a rocket motor whose leaking exhaust ignited the hydrogen fuel tank of the ill-fated *Challenger* space craft.

• In June1996 a satellite-lifting rocket *Ariane 5* turned cartwheels shortly after launch and scattered itself, a payload worth over half a billion dollars, and the hopes of European scientists over a marsh in French Guiana. The immediate cause was an integer oveflow exception in a subprogram whose result would later have been disregarded harmlessly. However, a punitive policy of treating all unanticipated exceptions as errors caused control to be wrested from the program guiding the rocket, thus punishing *it* instead of the programmers who blundered.

Most of us purchase insurance policies that will compensate us for parts of the costs of some of the rare calamities that may befall us. How much should we pay for that kind of insurance?

Besides allowing for the buyer's aversion to risk and the insurance company's overhead and profit, the cost of insurance should include the expected cost of the risk the company accepts; that expected cost equals the probability of the rare calamity multiplied by its cost if it occurs. Because the full cost of a calamity is usually too difficult to estimate in advance, the company limits its risk by agreeing to pay for losses up to a limit beyond which the insured will have to bear them himself. Moreover, the company must inflate its estimate of a calamity's probability by some amount, depending upon how capricious is the company's experience with similar risks, lest a run of bad luck deplete the company's resources, as happened recently to *Lloyd's of London*. Still, the probability of a very rare calamity can only be estimated, and different companies' actuaries sometimes come up with estimates that differ by an order of magnitude or more.

That a mathematically definable probability should be so much a matter of opinion may surprise someone who does not realize that probability is more than mere mathematics; it is more like physics or other sciences. (This is why universities now have Statistics departments separate from their Mathematics departments although, so recently as 1960, they were not separate at universities as respectable as U.C. Berkeley and Toronto.) Estimates of probabilities derive from models of situations and depend upon the fidelity of those models and the care with which details are taken into account. Estimates of probabilities of rare events are especially vulnerable to a methodological error:–

the disregard of correlations that would not matter
if the estimated probabilities were not so tiny.

After a calamity has been traced to a design error due to a misunderstanding or misconception, why should we believe that the design is free from other related errors, or that the known error's most likely consequences are fully comprehended?

> Shortly after Pan Am #103 blew up in 1988 over Lockerbie, Scotland, a passenger boarding an aircraft in Chicago was stopped when a bomb was found in his briefcase. He explained it thus: " I am terrified of being blown up on an aircraft. My travel agent tried to reassure me by explaining that the probability of a bomb on an aircraft is less than one in a million. If this is the case, the probability of two bombs is less than one in a trillion, so I figured I would be that much safer if I brought my own bomb." (This was a sick joke, not a true story.)
>
> . . .
>
> To lessen the DC-10's vulnerability to collisions, the conduits from the cockpit of this jumbo jet to its rear engine and tail control surfaces were duplicated and run along opposite edges of the passenger-compartment's floor. But early DC-10s had an ill-designed latch mechanism on their baggage-compartment door which consequently could not always withstand internal air pressurization at high altitudes. If the door blew out, the passenger-compartment's floor would collapse into the now depressurized baggage-compartment, rupturing both conduits and depriving the pilots of control over the rear engine and tailplanes. It happened to an American Airlines flight out of Detroit, but the pilot found a way to control the aircraft's attitude by adjusting the thrust of the under-wing engines and managed miraculously to land safely. It happened again to a Turkish Airlines flight out of Orly near Paris, France; this time the pilot found no way to avert a crash that killed all 400 on board.
>
> This calamity occurred partly because the two conduits' failure modes, though intended to be independent, were actually correlated by the floor to which both conduits were attached. The Lockheed 1011, a similar jumbo jet designed before the DC-10, had run a third conduit along the ceiling to avoid that correlation.

Our obligation is clear: We must assess the extent to which correlations initially deemed insignificant can vitiate an estimate of a rare calamity's probability if they are ignored.

Correlated Attributes

Let M and B be two attributes which every member of some population may or may not possess. Examples include …

Table 1: Two Attributes of Various Populations

Population:	Humans	Trapped Wolves	Terrorist Crimes	Computer Programs
Attribute M :	Male	Mature	Murder	Mistaken
Attribute B :	Bald	Banded	Bombing	Written in *BASIC*

We can also let $\mathbf{M}$ and $\mathbf{B}$ denote subsets of that population possessing the attributes, and let $\overline{\mathbf{M}}$ and $\overline{\mathbf{B}}$ denote the complementary subsets of members that lack the respective attributes. Now the whole population can be partitioned into four mutually exclusive subsets:

$$\mathbf{M}\cap\mathbf{B}\ ,\ \ \mathbf{M}\cap\overline{\mathbf{B}}\ ,\ \ \overline{\mathbf{M}}\cap\mathbf{B}\ ,\ \text{ and }\ \overline{\mathbf{M}}\cap\overline{\mathbf{B}}\ .$$

The respective counts of these subsets' members will be denoted by nonnegative variables:

$$\mathrm{mb} := |\mathbf{M}\cap\mathbf{B}|\ ,\qquad \mathrm{m}\overline{\mathrm{b}} := |\mathbf{M}\cap\overline{\mathbf{B}}|\ ,\qquad \overline{\mathrm{m}}\mathrm{b} := |\overline{\mathbf{M}}\cap\mathbf{B}|\quad \text{and}\quad \overline{\mathrm{mb}} := |\overline{\mathbf{M}}\cap\overline{\mathbf{B}}|\ .$$

Here " mb " and " $\mathrm{m}\overline{\mathrm{b}}$ " *etc.* are each a single symbol, not a product of two variables. And

$$\mu := |\mathbf{M}| = \mathrm{mb}+\mathrm{m}\overline{\mathrm{b}}\ ,\quad \overline{\mu} := |\overline{\mathbf{M}}| = \overline{\mathrm{m}}\mathrm{b}+\overline{\mathrm{mb}}\ ,\quad ß := |\mathbf{B}| = \mathrm{mb}+\overline{\mathrm{m}}\mathrm{b}\quad\text{and}\quad \overline{ß} := |\overline{\mathbf{B}}| = \mathrm{m}\overline{\mathrm{b}}+\overline{\mathrm{mb}}$$

are subtotals counting the members in the subsets originally described, though $\mathbf{M}$ may overlap $\mathbf{B}$ and $\overline{\mathbf{B}}$, *etc.* Finally total $n := \mu+\overline{\mu} = ß+\overline{ß} = \mathrm{mb}+\mathrm{m}\overline{\mathrm{b}}+\overline{\mathrm{m}}\mathrm{b}+\overline{\mathrm{mb}}$ counts the whole population. The relations among these numbers are summarized in …

Table 2: Counts of Subsets of a Population

Subsets	**B**	$\overline{\mathbf{B}}$	Subtotals
M :	mb	$\mathrm{m}\overline{\mathrm{b}}$	$\mu = \mathrm{mb}+\mathrm{m}\overline{\mathrm{b}}$
$\overline{\mathbf{M}}$:	$\overline{\mathrm{m}}\mathrm{b}$	$\overline{\mathrm{mb}}$	$\overline{\mu} = \overline{\mathrm{m}}\mathrm{b}+\overline{\mathrm{mb}}$
Subtotals:	$ß = \mathrm{mb}+\overline{\mathrm{m}}\mathrm{b}$	$\overline{ß} = \mathrm{m}\overline{\mathrm{b}}+\overline{\mathrm{mb}}$	$n = \mu+\overline{\mu} = ß+\overline{ß}$

Although the numerical entries in this table are all subsets' population counts, they can also be probabilities that a randomly selected member of the whole population belongs to a subset. To convert counts to probabilities, divide all counts by n or, what amounts to the same thing, set $n := 1$ and reinterpret the other variables as probabilities that a randomly selected member of the population possesses various attributes thus:

$\mathrm{mb} =$ Probability(M and B) , $\mathrm{m}\overline{\mathrm{b}} =$ Probability(M but not B) , … ,
$\mu =$ Probability(M) , $\overline{\mu} =$ Probability(not M) , … .

Except for replacing n by 1 , every algebraic relation among variables in Table 2 persists if they are reinterpreted as probabilities, but instead of "subtotals" the variables μ, $\overline{\mu}$, $ß$ and $\overline{ß}$ are called "Marginal Probabilities".

(Do not confuse *Marginal* Probabilities like Probability(M regardless of B) = μ with *Conditional* Probabilities like Probability(M | B) := Probability(M given B) = mb/ß .)

Given positive values for the four subtotals or marginal probabilities μ, $\overline{\mu}$, $ß$ and $\overline{ß}$, to what extent do the six equations in Table 2 determine the four counts or probabilities mb, $\mathrm{m}\overline{\mathrm{b}}$, $\overline{\mathrm{m}}\mathrm{b}$ and $\overline{\mathrm{mb}}$? These cannot be determined at all unless the

$$\text{Consistency Condition:}\quad n := \mu+\overline{\mu} = ß+\overline{ß}\ > 0$$

is satisfied, in which case the other four equations in Table 2 can be satisfied by infinitely many choices of mb, $m\bar{b}$, $\bar{m}b$ and $\overline{mb}$. To determine these four variables uniquely we have to know more about M and B .

For instance, the four variables are determined uniquely when attributes M and B are known to be *Statistically Independent*, which means $mb/n = (\mu/n)\cdot(ß/n)$ or, in other words,

Probability(M and B) = Probability(M regardless of B)·Probability(B regardless of M) .

In this case some algebra reveals first that $mb\cdot\overline{mb} - m\bar{b}\cdot\bar{m}b = 0$ and then that

$$mb = \mu\cdotß/n , \qquad m\bar{b} = \mu\cdot\bar{ß}/n , \qquad \bar{m}b = \bar{\mu}\cdotß/n \quad \text{and} \quad \overline{mb} = \bar{\mu}\cdot\bar{ß}/n .$$

But if M and B are *not* statistically independent, we have to know how they are *Correlated* in order to determine mb, $m\bar{b}$, $\bar{m}b$ and $\overline{mb}$ from μ, $\bar{\mu}$, ß and $\bar{ß}$. For this purpose we need the value of a *Correlation Coefficient* ç like the one defined thus:

$$\text{Correlation Coefficient} \quad ç := (mb\cdot\overline{mb} - m\bar{b}\cdot\bar{m}b)/\sqrt{(\mu\cdot\bar{\mu}\cdotß\cdot\bar{ß})} .$$

Degenerate situations shall be ruled out by the assumption $\mu\cdot\bar{\mu}\cdotß\cdot\bar{ß} > 0$, which means that none of **M** , $\overline{\mathbf{M}}$, **B** nor $\overline{\mathbf{B}}$ can be empty. Then it turns out that $-1 \le ç \le 1$ because (after a lot of algebraic manipulation)

$$1 - ç^2 = n\cdot(mb\cdot m\bar{b}\cdot\bar{m}b + mb\cdot m\bar{b}\cdot\overline{mb} + mb\cdot\bar{m}b\cdot\overline{mb} + \bar{m}b\cdot m\bar{b}\cdot\overline{mb})/(\mu\cdot\bar{\mu}\cdotß\cdot\bar{ß}) \ge 0 .$$

And $ç = 0$ just when M and B are statistically independent. $ç = 1$ just when $m\bar{b} = \bar{m}b = 0$, which means that M and B are so perfectly correlated that each implies the other. $ç = -1$ just when $mb = \overline{mb} = 0$, which means M and B are so perfectly anti-correlated that each precludes the other. Still, good reasons exist to wonder where this ç came from. Let's see … .

Suppose a part of whatever "causes" M may tend also to cause or inhibit B , and suppose ç is proportional to the strength of that partial cause. $ç = 0$ just when M and B are statistically independent; otherwise the counts or probabilities of $\mathbf{M}\cap\mathbf{B}$, $\mathbf{M}\cap\overline{\mathbf{B}}$, $\overline{\mathbf{M}}\cap\mathbf{B}$ and $\overline{\mathbf{M}}\cap\overline{\mathbf{B}}$ should depart by amounts proportional to ç from their values when $ç = 0$. After some algebra, the constants of proportionality turn out to be just one "constant" λ , say, that must satisfy

$$mb = (\mu\cdotß + \lambda\cdotç)/n , \quad m\bar{b} = (\mu\cdot\bar{ß} - \lambda\cdotç)/n , \quad \bar{m}b = (\bar{\mu}\cdotß - \lambda\cdotç)/n \quad \text{and} \quad \overline{mb} = (\bar{\mu}\cdot\bar{ß} + \lambda\cdotç)/n$$

in order to satisfy the equations in Table 2 above when $ç \ne 0$. Consequently, as expected,

$$ç = (mb\cdot\overline{mb} - m\bar{b}\cdot\bar{m}b)/\lambda .$$

The "constant" λ actually has to depend upon only μ, $\bar{\mu}$, ß and $\bar{ß}$ in such a way that $ç < 1$ unless $m\bar{b} = \bar{m}b = 0$, and $ç > -1$ unless $mb = \overline{mb} = 0$. These objectives are accomplished by $\lambda := \sqrt{(\mu\cdot\bar{\mu}\cdotß\cdot\bar{ß})}$ but other choices for λ might do as well for all we know. The following digression, which may be skipped on first reading, will justify our chosen λ geometrically.

How Closely does a Cloud Resemble a Line Segment ?

Consider two functions x and y that take values x_i and y_i for each individual i in a given population. If i will be chosen at random then x and y become *Random Variables* whose distributions depend upon the probability of choosing i ; assuming each individual as likely to be chosen as any other simplifies our explanation without weakening it. In the Cartesian (x, y)-plane, plot all the points with coordinates (x_i, y_i) . Does this cloud of points form a recognizable pattern, perhaps by clustering closely along some simple curve? If so, the more nearly the cloud conforms to such a pattern, the more strongly do we deem x and y to be correlated in the given population.

That a cloud of points may very nearly " form a recognizable pattern " suggests that the pattern has been seen before and, perhaps, given a name. Some people recognize more patterns than others; for example, one person may observe that nearly all of a cloud's points reside very near a spiral that another person has overlooked. This means that the recognition of a pattern comes about because the cloud has been subjected to a battery of tests, each designed to measure the closeness of a cloud to one of a known family of simple patterns, and has passed at least one test. One test may measure closeness to straight lines, another to conic sections like circles and ellipses and hyperbolas, another to checkerboards, another to logarithmic spirals, another to sinusoids, and so on. Each such test computes its own Correlation Coefficient to gauge how closely the closest member of its family comes to capturing all the points in the cloud.

Testing for straight line segments is easier than most other tests, yet it will be adequate for our purposes when applied to a population with only two attributes. The test appraises the closeness of the line segment closest to holding all the points (x_i, y_i) . The test must have five properties:

- It is independent of where the origin $(0, 0)$ lies amidst the cloud of points (x_i, y_i) .
- It is independent of the units (miles or millimeters, *etc.*) in which x and y are measured.
- It is independent of the number of points in the cloud provided they be not too few.
- It is unchanged if x and y are exchanged.
- It distinguishes correlation from anticorrelation; (x, y) and $(x, -y)$ are correlated oppositely.

To achieve the first two properties, the test standardizes x and y by taking account of …

$n := \sum_i 1$ = the number of individuals in the given population, and the *statistics*

$\bar{x} := \sum_i x_i/n$ = the *average*, *mean* or *expected* value of x in the given population,

$\sigma := \sqrt{\left(\sum_i (x_i-\bar{x})^2/n \right)}$ = the *standard deviation* of x in the given population,

$\bar{y} := \sum_i y_i/n$ = the average value of y in the given population, and

$\tau := \sqrt{\left(\sum_i (y_i-\bar{y})^2/n \right)}$ = the standard deviation of y in the given population.

Presumably $n > 2$ and $\sigma\cdot\tau > 0$ lest the test be futile. Later we'll need also the statistic

$\gamma := \sum_i (x_i-\bar{x})\cdot(y_i-\bar{y})/n$ = the *covariance* of x and y in the given population.

Now transform x into $\xi := (x-\bar{x})/\sigma$ and y into $\eta := (y-\bar{y})/\tau$ and obtain a transformed set of points (ξ_i, η_i) . These do not change if, before computing $\bar{x}$ and σ , an arbitrary constant C is added to all the x_i s , or if they are all multiplied by an arbitrary positive constant P ; and similarly for all the y_i s . This achieves the test's first two properties; the transformed cloud of points (ξ_i, η_i) exhibits the same pattern, if any, as does the original cloud of points (x_i, y_i) , but now standardized by shifts of origin and changes of scale to satisfy $\sum_i \xi_i = \sum_i \eta_i = 0$, so both avarages $\bar{\xi} = \bar{\eta} = 0$, and $\sum_i \xi_i^2/n = \sum_i \eta_i^2/n = 1$ so both standard deviations equal 1 .

Question: Of all straight lines in the Cartesian (ξ, η)-plane, which line £ minimizes the average of squared distances from £ to all points (ξ_i, η_i) , and how small is the minimized average?

High-school algebra will show how £ and the minimized average depend only upon covariance

$$ç := \sum_i \xi_i\cdot\eta_i/n = \gamma/(\sigma\cdot\tau) .$$

This *Product-Moment Correlation Coefficient* ç will serve as the gauge for our test of how nearly the given population's cloud of points (x_i, y_i) lies in a line segment. Later we'll see how ç is independent of n , as required by the third property above. The values of ç for all clouds sweep out the interval $-1 \le ç \le 1$, as can be confirmed by using Lagrange's identity:

$$0 \le \sum_i \sum_j (\xi_i \cdot \eta_j - \xi_j \cdot \eta_i)^2 = \ldots = 2n^2 \cdot (1 - ç^2) .$$

The inequality here becomes equality just when either every $\xi_i = \eta_i$ or else every $\xi_i = -\eta_i$. As random clouds go, the ones with ç near ± 1 are rather rare because all their points lie close to a line segment. The likeliest clouds have ç near zero; some of these look nonrandom but all show no substantial propensity towards one direction instead of another, as we shall see next.

Let $S^2(£)$ denote the average of all squares of distances from the points (ξ_i, η_i) to a line £ in the (ξ, η)-plane. The points (ξ, η) on £ satisfy an equation

$$\xi \cdot \sin(\theta) - \eta \cdot \cos(\theta) - \delta = 0$$

when £ is inclined at an angle θ to the ξ-axis and displaced a distance δ from $(0, 0)$. Then

$$S^2(£) = \sum_i (\xi_i \cdot \sin(\theta) - \eta_i \cdot \cos(\theta) - \delta)^2/n = \ldots = 1 - ç \cdot \sin(2\theta) + \delta^2 .$$

To minimize $S^2(£)$ we must first set $\delta = 0$; this means that only lines £ through $(0, 0)$ need be considered. Next we must choose £ 's inclination θ to minimize $S^2(£)$; the minimizing choice depends upon ç . If $ç = 0$ the inclination θ doesn't matter; $S^2(£) = 1$ for all lines £ through $(0, 0)$, signifying that x and y are uncorrelated (though not necessarily statistically independent). If $ç \ne 0$ lines £ of different inclinations θ have averaged squared distances $S^2(£)$ ranging from $1-ç$ to $1+ç$; the minimizing inclination is $\theta = \mathrm{sign}(ç) \cdot \pi/4$ and the minimized $S^2(£) = 1 - |ç|$, signifying that x and y are the more strongly (anti)correlated according as $(-)ç$ is closer to 1 . Thus, answering the Question above also explains ç 's significance: $|ç| = 1 -$ (the minimized average of squared distances) .

How is ç independent of n ? Actually, the larger the number n of points in a cloud, the more likely is its ç to be tiny; but that does not contradict ç 's independence of n . What " independence " means here is that if two clouds have the same ç but possibly different numbers n of points, and then if both clouds are standardized into the (ξ, η)-plane and combined, this union of two clouds is a cloud with the same correlation coefficient ç . The verification of this assertion, like the algebraic work implicit in each ellipsis "…" above, is left to the reader.

A simple example will illustrate how random variables x and y can be uncorrelated, in so far as $ç = 0$, but not statistically independent. Suppose the three points $(-1, -1)$, $(0, 2)$ and $(1, -1)$ in the (x, y)-plane have the same probability 1/3 . We find $\bar{x} = \bar{y} = 0$, $\sigma = \sqrt{2}$, $\tau = \sqrt{6}$ and $\gamma = 0 = ç$, so variables x and y are uncorrelated in this population. Their marginal probabilities are $\mathrm{prob}(x = -1) = \mathrm{prob}(x = 0) = \mathrm{prob}(x = 1) = \mathrm{prob}(y = 2) = 1/3$ and $\mathrm{prob}(y = -1) = 2/3$. But $1/3 = \mathrm{prob}(x = 0 \ \& \ y = 2) \ne \mathrm{prob}(x = 0) \cdot \mathrm{prob}(y = 2) = 1/9$; x and y aren't independent.

Next let us compute a correlation coefficient ç for two functions x and y defined, for every individual i in the population described by Table 2, as follows:

x_i takes one of two values, one if i has attribute M , otherwise the other value.

y_i takes one of two values, one if i has attribute B , otherwise the other value.

The two values for x do not matter so long as they are different; similarly for y . After a lot of algebra taking their averages, standard deviations and covariance into account, the result is

$$\text{Correlation Coefficient} \quad ç := (mb \cdot \overline{mb} - m\bar{b} \cdot \bar{m}b)/\sqrt{(\mu \cdot \bar{\mu} \cdot ß \cdot \bar{ß})} .$$

Now, knowing where this ç came from, we see that it is not arbitrary.

Correlation Neglected

Probabilistic inferences about rare events are too easily vitiated by small correlations that affect commonplace events negligibly. For instance, suppose a calamity occurs only when two rare events M and B both occur, and let their probabilities be μ and β respectively, both very tiny numbers like 10^{-6}. The calamity's probability $\mu\cdot\beta$ is tinier by far provided M and B are independent. But if M and B are slightly correlated, with small correlation coefficient ç, then the probability of calamity can turn out far bigger than $\mu\cdot\beta$, depending far more strongly upon ç than do the probabilities of other compound events like (M but not B). Here are the relevant formulas simplified by the substitutions $\bar{\mu} := 1-\mu$ and $\bar{\beta} := 1-\beta$:

Table 3: Probabilities of Compound Events

Compound Events	Probabilities	Constraints upon ç
M and B (calamity)	$mb = \mu\cdot\beta + \text{ç}\cdot\sqrt{(\mu\cdot\bar{\mu}\cdot\beta\cdot\bar{\beta})}$	Every Probability ≥ 0 , so ç must lie between $-\min_{\pm} (\mu\cdot\beta/(\bar{\mu}\cdot\bar{\beta}))^{\pm 1/2} \geq -1$ and $\min_{\pm} (\mu\cdot\bar{\beta}/(\bar{\mu}\cdot\beta))^{\pm 1/2} \leq 1$.
M but not B	$m\bar{b} = \mu\cdot\bar{\beta} - \text{ç}\cdot\sqrt{(\mu\cdot\bar{\mu}\cdot\beta\cdot\bar{\beta})}$	
B but not M	$\bar{m}b = \bar{\mu}\cdot\beta - \text{ç}\cdot\sqrt{(\mu\cdot\bar{\mu}\cdot\beta\cdot\bar{\beta})}$	
Neither B nor M	$\overline{mb} = \bar{\mu}\cdot\bar{\beta} + \text{ç}\cdot\sqrt{(\mu\cdot\bar{\mu}\cdot\beta\cdot\bar{\beta})}$	

For example take $\mu = 10^{-6}$, $\beta = 4\cdot 10^{-6}$, and consider the effects of three values

$\text{ç} = 10^{-2}$ (M slightly correlated with B), or
$= 0$ (M independent of B), or
$= -2\cdot 10^{-6}$ (M very slightly anti-correlated with B).

All three of these correlations would be considered small, perhaps small enough to be ignored, and yet their influence upon the probability of calamity changes it by orders of magnitude:

Table 4: Probabilities of Correlated Events

Compound Events	$\text{ç} = 10^{-2}$	$\text{ç} = 0$	$c = -2\cdot 10^{-6}$
M and B (calamity)	$2\cdot 10^{-8}$	$4\cdot 10^{-12}$	10^{-17}
M but not B	$9.80\cdot 10^{-7}$	$1.00\cdot 10^{-6}$	$1.00\cdot 10^{-6}$
B but not M	$3.98\cdot 10^{-6}$	$4.00\cdot 10^{-6}$	$4.00\cdot 10^{-6}$
Neither B nor M	0.999995	0.999995	0.999995

Thus can probabilistic assessments of rare events be invalidated utterly by almost imperceptible correlations that are customarily ignored quite properly during assessments of populations' gross central tendencies. The tinier the correlations, the harder they are to ascertain; but to ignore them merely because they are tiny is methodologically unsound and sometimes dangerous.

If all we know about M and B is their respective probabilities μ and β ,
if we are ignorant about how M and B influence each other, then for all we know
the probability of a calamitous coincidence (M and B) could be as big as $\min\{\mu, \beta\}$,
not the far tinier estimate $\mu\cdot\beta$ of calamity's probability if M and B are independent.