has the form

$$\pm(n-d)\alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_d},$$

where $1 \le i_1 < i_2 < \cdots < i_d \le n$. Indeed, each positive monomial $\alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_d}$ in $s_{d+1}\tilde{P}_1$ occurs exactly as a product $(\alpha_{i_1}\cdots\alpha_{i_{k-1}}\alpha_j\alpha_{i_k}\cdots\alpha_{i_d})(1/\alpha_j)$, where the index j belongs to $\{1, 2, \ldots, n\} \setminus \{i_1, \ldots, i_d\}$ and $i_{k-1} < j < i_k$. (Of course if $j < i_1$, then α_j is the first factor, while if $i_d < j$, then α_j is the last factor in the bracket.) Therefore each such summand $\alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_d}$ occurs $n-d$ times. Precisely because $(n-d)s_d$ consists of the monomials $\mp(n-d)\alpha_{i_1}\cdots\alpha_{i_d}$ where $1 \le i_1 < i_2 < \cdots < i_d \le n$, we see that B contains no positive monomial.

The upshot: we see that in the sum $A+B$ there are only negative monomials, whereas in C there are only positive monomials. Therefore no monomial in $A+B$ can be cancelled out by a monomial in C. Thus $A+B=0=C$, and our proof is complete.

In conclusion, we emphasize that the foregoing proof is both transparent and natural. It extends the straightforward proof in case 1, and it demonstrates that the same idea works nicely, provided one is not afraid of negative exponents.

ACKNOWLEDGMENTS. I would like to thank Zinovy Reichstein for our many interesting discussions, which at one moment involved this particular proof. I would also like to thank my enthusiastic Galois theory students in Math. 403b, who inspired me once again to reflect on the basic principles of Galois theory. And last but not least, I would like to thank the referee of this note for many constructive insights and help in improving the exposition of same. This work was partially supported by NSERC grant R0370A01.

REFERENCES

1. N. Bourbaki, *Algèbre*, Masson, Paris, 1981.
2. J.-P. Escofier, *Galois Theory*, Springer-Verlag, New York, 2001.
3. D. Mead, Newton's identities, this MONTHLY **99** (1992) 749–751.
4. P. Morandi, *Field and Galois Theory*, Springer-Verlag, New York, 1996.
5. Z. Reichstein, An inductive proof of Newton's identities, *Matematicheskoe Prosveschenie* [Mathematics Education] **3** (2000) 204–205 (Russian).
6. D. Zeilberger, A combinatorial proof of Newton's identities, *Discrete Math.* **49** (1984) 319.

The University of Western Ontario, Department of Mathematics, Middlesex College, London, Ontario, Canada N6A 5B7
minac@uwo.ca

An Illuminating Counterexample

Michael Hardy

Suppose that $X_1, \ldots, X_n$ are independent random variables with a normal (or "Gaussian") distribution with expectation μ and variance σ^2. A statistician who has observed the values of $X_1, \ldots, X_n$ must guess the values of μ and σ^2. Among the statistically naive, it is sometimes asserted that

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2,$$

where $\overline{X} = (X_1 + \cdots + X_n)/n$, is a better estimator of σ^2 than is

$$T^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2,$$

because S^2 is "unbiased" and T^2 is "biased." That S^2 is unbiased means $E(S^2) = \sigma^2$, i.e., an "unbiased estimator" is a statistic whose expected value is the quantity to be estimated.

The goodness of an estimator is sometimes measured by the smallness of its "mean squared error," defined as $E(([\text{estimator}] - [\text{quantity to be estimated}])^2)$. By that criterion the biased estimator T^2 would be better than the unbiased estimator S^2, since

$$E((T^2 - \sigma^2)^2) < E((S^2 - \sigma^2)^2),$$

but the difference is so slight that no one's statistical conscience is horrified by anyone's preferring S^2 over T^2. Besides, the smallness of the mean squared error as a criterion for evaluating estimators is not necessarily sacred anyway.

A more damning example, well-known among statisticians, is described in [**1**, p. 168]. We have $X \sim \text{Poisson}(\lambda)$, so that $P(X = x) = \lambda^x e^{-\lambda}/x!$ for $x = 0, 1, 2, \ldots,$ and $P(X = 0)^2 = e^{-2\lambda}$ is to be estimated. Any unbiased estimator $\delta(X)$ satisfies

$$E(\delta(X)) = \sum_{x=0}^{\infty}\delta(x)\frac{\lambda^x\, e^{-\lambda}}{x!} = e^{-2\lambda}$$

uniformly in $\lambda \geq 0$. Clearly the only such function is $\delta(x) = (-1)^x$. Thus, if it is observed that $X = 200$, so that it is astronomically implausible that $e^{-2\lambda}$ is anywhere near 1, the desideratum of unbiasedness nonetheless requires us to use $(-1)^{200} = 1$ as our estimate of $e^{-2\lambda}$. And if $X = 3$ is observed, the situation is even more absurd: we must use $(-1)^3 = -1$ as an estimate of a quantity that we know to be in the interval $(0, 1]$. A far better estimator of $e^{-2\lambda}$ is the biased estimator e^{-2X} (which is the answer given by the well-known method of maximum likelihood).

Here is a different counterexample, which the visually inclined may find even more horrifying. A light source is at an unknown location μ somewhere in the disk $D = \{(x, y) : x^2 + y^2 \leq 1\}$ in the Euclidean plane (see Figure 1). A dart thrown at

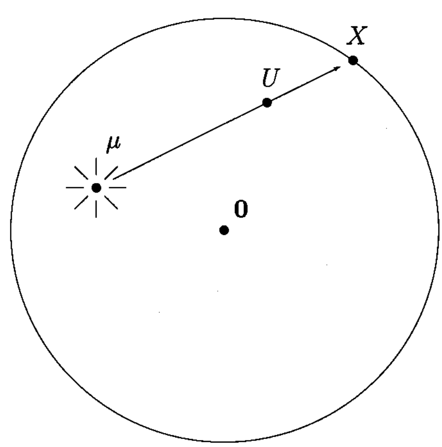

Figure 1. $D = \{(x, y) : x^2 + y^2 = 1\}$.

the disk strikes some random location U in the disk, casting a shadow at a point X on the boundary. The random variable U is uniformly distributed in the disk, i.e., the probability that it is within any particular region is proportional to the area of the region. The boundary is a translucent screen, so that an observer located outside of the disk can see the location X of the shadow, but cannot see where either the light source or the opaque object is. Given only that information—the location X of the shadow—the location μ of the light source must be guessed.

A common-sense approach to guessing μ might proceed as follows: *Before* we observe the shadow, our information is invariant under rotations, and so should be our estimate. Therefore, we use 0 in $\mathbb{R}^2$ as our prior (i.e., pre-data) estimate. Then, when we observe X, since X is more likely to be far from the light source than close to it, we adjust our estimate by moving it away from the shadow. Because the amount of information in the shadow is small, we don't move it very far. We get an estimator of the form cX with $c < 0$, but c is not very much less than 0.

If we insist on unbiasedness, we must choose c so that $E(cX) = \mu$ uniformly in μ. To think about that, we first express the problem in polar coordinates. Write $\mu = \rho(\cos\varphi, \sin\varphi)$ and $X = (\cos\Theta, \sin\Theta)$.

Proposition. *The probability distribution of the random angle Θ is given by*

$$P(d\theta) = \frac{1 - \rho\cos(\theta - \varphi)}{2\pi}\, d\theta. \tag{1}$$

From this proposition it follows that $E(X) = -\mu/2$. Therefore, our unbiased estimator is $cX = -2X$, which is always absurdly remote from the disk, by a full radius!

Proof of the proposition. A simplification will follow from the observation that the way in which the probability distribution $P(d\theta)$ depends on μ is both rotation-equivariant and affine. That it is affine means that if the probability distribution of Θ is $P_\mu(d\theta)$ when the light source is at μ then $P_{a\mu+(1-a)\nu}(d\theta) = aP_\mu(d\theta) + (1-a)P_\nu(d\theta)$ for any value of a for which $a\mu + (1-a)\nu$ remains within the disk. (An affine mapping is one that preserves linear combinations in which the sum of the coefficients is 1; a linear combination satisfying that constraint is an "affine combination.") To see that this mapping is affine, consider Figure 2. The area between μ and the arc from

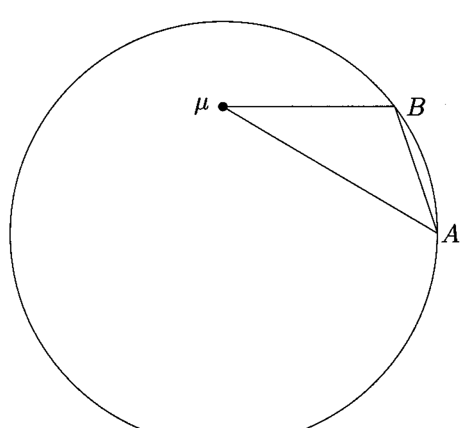

Figure 2.

A to B is the sum of the area of the triangle μAB and the area of the region bounded by the arc AB and the secant line AB. As μ moves, the area bounded by the arc and the secant line remains constant and the area of the triangle depends on μ in an affine fashion. The desired "affinity" follows.

Rotation-equivariance reduces the problem to finding the probability distribution when μ is between $(0, 0)$ and $(1, 0)$. "Affinity" reduces it from there to the problem of finding the probability distribution when μ is at either of those two points.

If $\mu = (0, 0)$, the probability distribution of Θ is clearly uniform on the interval from 0 to 2π, i.e., it is $d\theta/(2\pi)$. If $\mu = (1, 0)$, then for $0 \le \theta \le 2\pi$ we have

$$\begin{aligned} P(0 \le \Theta \le \theta) &= \frac{\text{area between arc and straight line from } (1, 0) \text{ to } (\cos\theta, \sin\theta)}{\text{area of disk}} \\ &= \frac{\theta - \sin\theta}{2\pi}. \end{aligned}$$

Differentiation yields

$$P(d\theta) = \frac{1 - \cos\theta}{2\pi}\, d\theta.$$

If $\mu = (\rho, 0)$, then by "affinity" we have

$$P(d\theta) = (1 - \rho)\frac{d\theta}{2\pi} + \rho\frac{(1 - \cos\theta)d\theta}{2\pi} = \frac{1 - \rho\cos\theta}{2\pi}\, d\theta.$$

Rotation-equivariance then gives (1). ∎

The Bayesian approach to statistical inference assigns probabilities not to *events* that are *random* (according to their relative frequencies of occurrence), but to *propositions* that are *uncertain* (according to the degree to which known evidence supports them). Accordingly, we can regard the location μ of the light source as uniformly distributed in the disk, and then use the conditional expected location $E(\mu|X)$ as an estimator of μ. Equation (1) gives the conditional distribution of Θ given μ; the marginal (i.e., "unconditional") distribution of $\mu = \rho(\cos\varphi, \sin\varphi)$ is given by

$$\frac{\rho\, d\rho\, d\varphi}{\pi}. \tag{2}$$

The joint distribution of (μ, Θ) is the product of (1) and (2):

$$\frac{(1 - \rho\cos(\theta - \varphi))\rho\, d\rho\, d\varphi\, d\theta}{2\pi^2}. \tag{3}$$

The conditional distribution of $\mu = \rho(\cos\varphi, \sin\varphi)$ given that $\Theta = \theta$ comes from regarding (3) as a function ρ and φ with θ fixed and normalizing:

$$P(d\rho, d\varphi|\Theta = \theta) = \frac{(1 - \rho\cos(\theta - \varphi))\rho\, d\rho\, d\varphi}{\text{constant}}.$$

Integration shows that the "constant" is π. Finally, we get

$$E(\mu|X) = \int_0^{2\pi} \int_0^1 \rho(\cos\varphi, \sin\varphi) \frac{1 - \rho\cos(\Theta - \varphi)}{\pi} \rho \, d\rho \, d\varphi$$
$$= -(\cos\Theta, \sin\Theta)/4 = -X/4,$$

which is an eminently reasonable estimator under the circumstances.

REFERENCE

1. J. P. Romano and A. F. Siegel, *Counterexamples in Probability and Statistics*, Wadsworth & Brooks/Cole, Monterey, CA, 1986.

Department of Mathematics, University of Toledo, 2801 W. Bancroft, Toledo OH 43606-3390
michael.hardy@utoledo.edu

A Norm Inequality for Hermitian Operators

Ritsuo Nakamoto

It is well known that the following inequality holds: for any real number θ

$$|e^{i\theta} - 1| \leq |\theta|; \tag{1}$$

equivalently,

$$|e^{i\theta_1} - e^{i\theta_2}| \leq |\theta_1 - \theta_2| \tag{2}$$

for all real numbers θ_1 and θ_2.

Inequality (1) is easily proved by a direct calculation. Geometrically, this is interpreted as a relationship between chordal distance and arclength on the unit circle in the complex plane.

Now, by the spectral theorem (see [**2**]), we have

$$\|e^{iH} - 1\| \leq \|H\|$$

for any Hermitian (bounded linear) operator H on a Hilbert space. Furthermore, if Hermitian operators H and K commute, we have an analogue of (2):

$$\|e^{iH} - e^{iK}\| \leq \|H - K\|. \tag{3}$$

It might be expected that (3) would hold for Hermitian operators H and K without the commutativity assumption.

On the other hand, we recall the differential equation $dX/dt = HX + XK$ for matrices. Its solution is given by $X = e^{tH}Ce^{tK}$, where $C = X(0)$ (see [**1**, chap. 10, Theorem 5]). Related to this, we note Lyapunov's theorem on stability, which says that the real part of each eigenvalue of a real matrix A is negative if and only if there exists a unique positive definite matrix Y satisfying $A'Y + YA = -I$, in which A' signifies the transpose of A (see [**1**, chap. 13, Theorem 2]).