A Good Proof is a Master Key

Debrabata Basu

In a hotel there are many rooms each with its individual key. Each key is made so crooked that it can open only one door lock. Only the hotel manager has the master key that is simple enough to turn inside all the locks. A good proof is such a master key. It is simple and straightforward and can unlock many a mathematical problem. Here is an example.

Let X be a real random variable with c.d.f. F. When and how does it follow that $F(X) \sim U$ (the uniform distribution over the unit interval)? It is my experience that, when this simple question is posed before a class of advanced graduate students, I get a surprising variety of crooked solutions. The master key to this problem is

(A) Whatever the c.d.f. F, the two events (i) $X \leq x$ and (ii) $F(X) \leq F(x)$ are identical for any fixed x.

Since F is monotone non-decreasing, (i) implies (ii) logically. In order to understand the reverse implication, consider the set $I_x = \{x^{'} : F(x^{'}) = F(x)\}$. Since F is non-decreasing, I_x has to be an interval containing x. [If F is strictly increasing at x, then I_x is the single-point set $\{x\}$ and in this case clearly (ii) implies (i).] Since F is everywhere right-continuous, the interval I_x is always closed at its left end. The right end-point b_x belongs to I_x if and only if F is continuous at b_x. If $b_x \in I_x$ then (ii) logically implies the event $X \leq b_x$ and this in turn implies the event $X \leq x$ since the interval $(x, b_x]$ has $F-$measure zero. On the other hand, if $b_x \notin I_x$ then we reach the same conclusion after replacing $X \leq b_x$ by $X < b_x$ and $(x, b_x]$ by (x, b_x) in the above argument. Thus (A) is always true. It imediately follow that

(B) $Prob[F(X) \leq F(x)] = F(x)$ for all x.

If F is continuous, then F(x) takes all values in the interval $(0, 1)$ and so $F(X) \sim U$.

I have found many students struggling hard through text-book proofs of Kolmogorov's famous result that with n i.i.d. random variables $X_1, X_2, \ldots, X_n$ with common continuous c.d.f. F, the quantity

D = sup | $\hat{\mathrm{F}}_n(x)$ - F(x) |

(where $\hat{\mathrm{F}}_n(x)$ is the empirical c.d.f.) is distribution-free. A quick turn of the master key (A) easily unlocks the problem. Writing $I(E)$ for the indicator of the event E, we can rewrite D as

$$D = sup_x|\frac{1}{n}\sum I(X_j \leq x) - F(x)|$$

$$= sup_x|\frac{1}{n}\sum I[F(X_j) \leq F(x)] - F(x)|$$

$$= sup_y|\frac{1}{n}\sum I(Y_j \leq y) - y|$$

where $y \in (0,1)$ and $Y_1, Y_2, \ldots, Y_n$ are i.i.d. U.

Making a right-left inversion of (B) we have

(C) If T is a real random variable with $G(t) = Prob(T \geq t)$, then $Prob[G(T) \leq G(t)] = G(t)$ for all t.

For a crooked verstion of the above simple master key refer to theorem (9.1) on p 224 of Kempthorn & Folks (1971) - the authors call this the fundamental theorem of test of significance. The context is as follows. Let H_0 be a simples null-hypothesis to be tested and let $T = T(X)$ be the test criterion - the larger the observed value of T the more significant is the test. For an observed value t of T, let us call $G(t) = Prob(T \geq t|H_0)$ the attained level of significance. R.A. Fisher noted that if we consistantly follow (in repeated trials) the rule "reject the null hypothesis if the attained level of significance is $\leq \alpha_0$", then the probability of first kind of error associated wih this rule is α_0. My advanced graduate students invariably get greatly puzzled when I ask them to prove Fisher's statement. Is the statement always true? The answer is yes provided α_0 is an attainable level of significance. Thus (C) is the so-called fundamental theorem of test of significance.

REFERENCE

Kempthorne, Oscar and Folks, Leroy (1971): Probability Statistics and Data Analysis. Ames, Iowa: The Iowa State University Press.