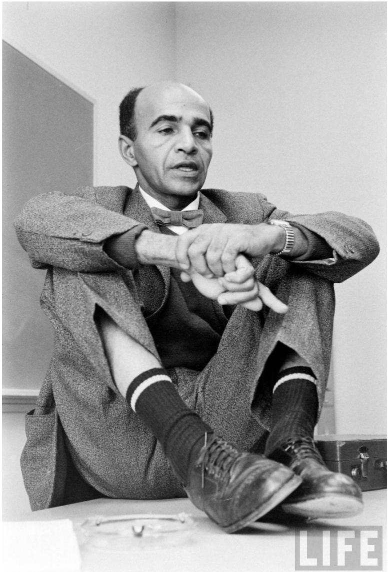
David H. Blackwell (1919-2010)



List of Early African American PhDs

Year	Name	University	Topic
1925	Elbert Frank Cox	Cornell	Differential Equ
1928	Dudley Weldon Woodard	UPenn	Topology
1933	William Schieffelin Claytor	UPenn	Topology
1934	Walter Richard Talbot	Pitt	Group Theory
1938	Reuben Roosevelt McDaniel	Cornell	Algebraic Number Theory
1938	Joseph Alphonso Pierce	Michigan	Statistics
1941	David Harold Blackwell	Illinois	Probability
1942	Jesse Ernest Wilkins	Chicago	Calculus of Variations
1943	M. Euphemia Lofton Haynes	Catholic University of America	Geometry

Blackwell's Favorite Papers

1. On a Class of Probability Spaces, **Berkeley Symposium on Mathematical Statistics and Probability**, 1956.
2. On Optimal Systems, **Annals of Mathematical Statistics**, 1954.

   On multi-component attrition games, **Naval Research Logistics**, 1954.

   An Analog of the Minimax Theorem for Vector Payoff, **Pacific Journal of Mathematics**, 1956.
3. Merging of Opinions with Increasing Information, **Annals of The Annals of Mathematical Statistics**, 1962 (with Lester Dubins).

Approachability

- $\{X_t\}$ sequence of bounded vectors
- $A_n = \frac{\sum_{t \leq n} X_t}{n}$
- $A_n \bullet X_{n+1} \leq 0$
- $\Rightarrow d(A_n, 0) \rightarrow 0$

Approachability

- $\{X_t\}$ sequence of bounded vectors, T a convex set (TARGET SET)
- $A_n = n^{-1} \sum_{t \le n} X_t$
- $\Pi_T(A_n)$ closest point on T to A_n
- $[A_n - \Pi_T(A_n)] \bullet [X_{n+1} - \Pi_T(A_n)] \le 0$
- $\Rightarrow d(A_n, T) \to 0$

Approachability

- $\{X_t\}$ sequence of bounded vectors, T a convex set (TARGET SET)
- $A_n = \frac{\sum_{t \leq n} X_t}{n}$
- $\Pi_T(A_n)$ closest point on T to A_n
- $E\{[A_n - \Pi_T(A_n)] \bullet [X_{n+1} - \Pi_T(A_n)] | A_n\} \leq 0$
- $\Rightarrow d(A_n, T) \to 0$ almost surely

Approachability

- Row (R) and Column (C) repeatedly meet to play a $m \times n$ matrix game.
- If R chooses i and C chooses j, the outcome is a vector v_{ij} in some compact space.
- Strategy chosen by R in round t is i_t
- Strategy chosen by C in round t is j_t
- $A_n = \sum_{t=1}^{n} v_{i_t j_t}/n$.
- Target set T (convex)

GOAL:

R must force (APPROACHABILITY) $d(A_n, T) \to 0$ almost surely as $n \to \infty$.

A convex set T is approachable if and only if every tangent hyperplane of T is approachable.

Extensions

Finite action apace	Compact action space
Euclidean norm	Other norms
Simple Average	Weighted Average
Finite Dimensions	Infinite Dimensions
Convex target set	Closed target set
Randomization	Deterministic
Full observability	Partial Observability
Discrete Time	Continuous Time

No Regret

Repeatedly predict the next element of an infinite sequence of 0's and 1's.

Measure the fraction, F_n, of INCORRECT guesses after n rounds.

- $H_n =$ fraction of 1's
- $1 - H_n =$ fraction of 0's
- $F_n \leq \min\{H_n, 1 - H_n\} + \epsilon_n$
- $\epsilon_n \rightarrow 0$ almost surely

No Regret

- $L_t^0 \in \{0, 1\}$ error at time t if predict 0
- $L_t^1 \in \{0, 1\}$ error at time t if predict 1
- Outcome from predicting 0 at time t is $(L_t^0 - L_t^1, 0)$
- Outcome from predicting 1 at time t is $(0, L_t^1 - L_t^0)$

No Regret

Let w_t be probability of predicting 0 at time t

$$A_n = \left(\frac{\sum_{t=1}^{n} w_t[L_t^0 - L_t^1]}{n}, \frac{\sum_{t=1}^{n}(1-w_t)[L_t^1 - L_t^0]}{n}\right)$$

Target Set = Non-positive Orthant

No Regret

From Approachability:

$$\frac{\sum_{t=1}^{n}(1-w_t)[L_t^1 - L_t^0]}{n} \leq \epsilon_n$$

$$\Rightarrow \frac{\sum_{t=1}^{n}(1-w_t)[L_t^1 - L_t^0]}{n} + \frac{\sum_{t=1}^{n} L_t^0}{n} \leq \frac{\sum_{t=1}^{n} L_t^0}{n} + \epsilon_n$$

$$\frac{\sum_{t=1}^{n}[w_t L_t^0 + (1-w_t)L_t^1]}{n} \leq \frac{\sum_{t=1}^{n} L_t^0}{n} + \epsilon_n = H_n + \epsilon_n$$

Similarly

$$\frac{\sum_{t=1}^{n}[w_t L_t^0 + (1-w_t)L_t^1]}{n} \leq \frac{\sum_{t=1}^{n} L_t^1}{n} + \epsilon_n = 1 - H_n + \epsilon_n$$

No External Regret

$L_i(n)$ = loss from using action i in period n (bounded).

$\exists$ randomized rule that incurs loss $R(n)$ in period n such that

$$\frac{\sum_{n=1}^{t} R(n)}{t} \leq \min_i \left\{ \frac{\sum_{n=1}^{t} L_i(n)}{t} \right\} + \epsilon_t$$

$\epsilon_t \to 0$ almost surely as $t \to \infty$.

Internal Regret

Look back at all times played strategy i.

What if, on *all* those occasions we had played strategy j instead.

If we would be better off by doing this, then we suffer (internal) regret.

Goal: play so as to *avoid* internal regret.

Internal Regret

1. $w_i(t)$ = probability of playing i at time t.
2. Expected loss upto time T is

$$\sum_{t=1}^{T}\sum_{k} w_k(t)L_k(t).$$

3. If, whenever action i was played we play j instead, payoff is

$$\sum_{i=1}^{T}\sum_{i} w_i(t)L_i(t) + \sum_{t=1}^{T} w_i(t)(L_j(t) - L_i(t)).$$

Internal Regret

Internal regret at time T is

$$R_T = \frac{\min_{i,j}[\sum_{t=1}^{T} w_i(t)(L_j(t) - L_i(t))]^-}{T}.$$

Goal is to choose $\{w(t)\}$'s using history so that

$$R_T \to 0.$$

Calibration

Each period you announce the probability p that the next term in a infinite 0-1 sequence will be a 1.

Announce different p's in different periods.

How should one measure accuracy of forecast?

Calibration

Compare average of the p's and average of the 1s.

outcome	0	1	0	1	0
p	0.5	0.5	0.5	0.5	0.5

Calibration

Break sequence up into two subsequences.

One corresponding to even periods and the other to odd periods.

On each subsequence look at average of the p's and average of the 1s.

outcome	0	0	1	1	0	0	1	1	0
p	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Calibration

In general would like to break up into different subsequences.

A checking rule is an algorithm that builds a subsequence period by period.

The decision to include a period can depend on the history of outcomes and forecasts to date.

To PASS a checking rule means that on the relevant subsequence the average of your p's is close to the average of the 1s.

Calibration

Naive Calibration: Restricted family of checking rules associated with forecasts made.

$\rho_t(p) =$ proportion of 1's in all periods upto t that forecast was p.

For all p announced sufficiently frequently

$$|\rho_t(p) - p| \to 0.$$

Calibration

1. Given a checking rule is there a forecasting algorithm that will pass it (for all possible realizations)?
2. Given a finite collection of checking rules is there a forecasting procedure that will pass ALL of them?
3. Countable collection?
4. ALL Checking rules?

Calibration

Highlights importance of how one measures accuracy of a prior/probability forecast.

Backward Guarantee: From some point in future, looking BACK, forecast close to realized outcome e.g. calibration

Forward Guarantee: From some point in future, looking FORWARD, forecast close to realized outcome e.g. merging.

Testing

- $S = \{0, 1\}$ = state space.
- S^n = all n length 0-1 sequences.
- S^* = all infinite 0-1 sequences.
- $s = (s_1, s_2, \ldots, s_n) \in S^n$ an n-sequence of outcomes.
- s_i = state realized in period i.
- Element of $[0, 1]$ is called a *forecast* of the event '1'.
- A forecast made at period r refers to outcomes that will be observed in period $r + 1$.
- Δ^* the set of probability distributions over $[0, 1]$.

Testing

- A forecasting algorithm is a function

$$f : \bigcup_{r=0}^{n-1} (S^{r+1} \times [0,1]^r) \to \Delta^*$$

- A (finite) *test* is a function $T : S^n \times [0,1]^n \to \{0,1\}$.

Testing

Denote unknown process that generates s by μ.

Test T is said to *pass the truth* with probability $1 - \epsilon$ if for all $s \in S^n$

$$\Pr_{\mu}(\{T(s, P(s)) = 1\}) \geq 1 - \epsilon.$$

Type I error (rejecting the null when it is true) is small.

Testing

T can be *ignorantly* passed by f with probability $1-\epsilon$ if for every $s \in S^n$,

$$\Pr_f(\{T(s, f(s)) = 1\}) \geq 1 - \epsilon.$$

Every finite test T that passes the truth *whp* can be ignorantly passed by some forecasting algorithm.