

АБРАХАМ ВАЛД И СТАТИСТИЧЕСКИТЕ ИЗСЛЕДВАНИЯ ВЪРХУ САМОЛЕТНИТЕ БРОНИ

¹Лъчезар Томов, ²Георги Дойчинов

¹ Нов български университет

² Национална гимназия за древни езици и култури „Св. Константин-Кирил Философ“

Резюме. Настоящата работа е посветена на изследванията на Абрахам Валд върху вероятността за оцеляване на самолет с даден брой поражения и на фолклорните легенди за неговата роля. Ние даваме ново обяснение на извеждането на базовото уравнение на Абрахам Валд и на основанията му за избор на допускания, които правят това извеждане възможно. Описваме подробно процеса на извеждане, за да осветим ролята на интуицията при статистическите изследвания и кога не бива да се доверява ученият на нея. Обръщаме внимание на ролята на оценките на интервалите при изчисляване на приблизителни стойности за липсващи данни. Осветляваме ролята на допусканията в процеса на научно изследване и методологията за тяхното добавяне и постепенно смекчаване.

Ключови думи: история; Валд; статистика; евристика; вероятности; епистемология

Абрахам Валд

Математикът Абрахам Валд е роден на 31 октомври 1902 г. в румънския град Клуж-Напока, който тогава се намирал в Австро-Унгарската империя. Той не получил формално образование и бил образован у дома от своите родители. Причината за това била, че семейството било силно религиозно, а според еврейската традиция събота, който бил официален учебен ден тогава, бил ден за почивка. Въпреки това младежът получил солидно образование. През 1928 г. завършил Клужкия университет „Бабеш-Болай“ (тогава университет „Фердинанд I“) с математика, а през 1931 г. защитил докторска степен по математика във Виенския университет. Заради дискриминацията срещу евреите в междувоенна Австрия не могъл да стане университетски преподавател. През 1938 г. Германия анексирала Австрия и математикът бил принуден да бяга. Установил се със семейството си в САЩ и започнал работа в комисията „Колс“ за изследване на икономиката. През Втората световна война приложил своите теоретични познания за статистиката на практика.

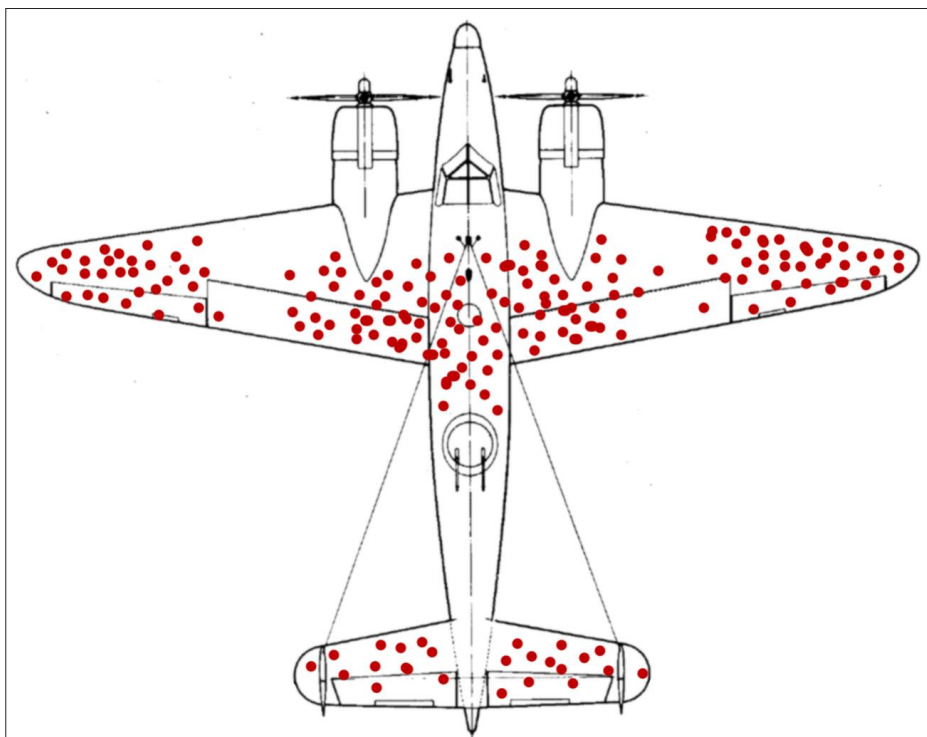
През 1950 г. бил поканен от правителството на Индия да проведе няколко лекции в страната. Той посетил Индийския статистически институт в Калку-

та и планирал да отиде на Индийския научен конгрес в Бангалор през януари 1951 г., но загинал в самолетна катастрофа над горския масив Нилигри в Южна Индия на 13 декември 1950 г.

Легендата разказва:

Тъй като за военовъздушните сили на САЩ било важно да установят къде да добавят допълнително броня на своите самолети, за да намалят загубите, а не можели да добавят навсякъде, тъй като било скъпо и неефективно, те се обърнали към Валд. Валд приложил метода на така наречения *Survivorship bias*, който е логическата грешка да се разглеждат случаите на преминалите през дадена селекция и на тяхна основа да се изготвя статистиката, а неуспелите да минат през селекцията да не се разглеждат, тъй като обикновено не е ясна причината за техния неуспех и не могат да бъдат забелязани.

Когато самолети се завръщали от бойна мисия (фиг.1), Валд проучвал къде са ударени от снаряди. Той вярвал, че вражеските артилеристи стрелят на случаен принцип.



Фигура 1. Местата, на които са уцелени завърналите се самолети¹⁾

Ако дадено място е било тежко засегнато, той смятал, че на това място не трябва да се поставя броня, тъй като щом самолетът се е завърнал, то няма нужда там да се поставя броня. Където обаче нямало следи от снаряди, Валд препоръчвал да се сложи броня, тъй като, ако това място е било засегнато, самолетът е щял да бъде свален. Свалените самолети не се проучвали, тъй като не били налични за изследване – намирали се на вражеска територия или под вода. Изследването на Валд се оказало ключово и след него загубите на самолети били намалени. Така Валд превърнал грешката в нещо полезно.

Всъщност, както (Mangel and Samaniego, 1984), предадено от (Bil Casselman, 2016) показва, истината е малко по-различна. Валд изобретява т.нар. секвенционен анализ (sequential analysis) и спомага за подобряване качеството на производството на американско оръжие, което е голям негов принос в статистиката и войната. Изследването, което той прави върху завърналите се американски самолети, е малко по-различно от предаденото в легендата. Той търси да изчисли вероятностите един самолет да оцелее, след като вече е бил ударен един, два или повече пъти. За всеки брой попадения вероятността е различна. Проблемът, който Валд решава, е как да се изчислят граници за тези вероятности, при положение че нямаме никакви данни за самолетите, които *не са се върнали*. Той намира начин да оцени щетите по самолетите, които не са се върнали – изключителна магия от теория на вероятностите.

Други прояви на евристика на наличността

Този принцип (*според фолклора*) бил приложен и през Първата световна война, когато полевите болници и лазарети били залети от британски войници с рани по главата, носещи шлемове Броди. Командването мислело да смени шлемовете с други, но статистиците го посъветвали да не го прави, тъй като фактът, че са стигнали до болницата, а не са били убити на място, показва, че шлемовете са ефективни. Ако от 10 простреляни в главата умират 9 без шлем, а с шлем умира 1 и са ранени 6 души, това показва, че шлемът работи, но болниците се пълнят с ранени войници. Заради това тези шлемове останали на служба и се използвали до началото на Втората световна война.

Въпреки ползите от употребата му не бива да се забравя, че Survivorship bias е логическа грешка. Антиваксърите (хората, които са против ваксините) твърдят, че няма нужда и смисъл от задължителните ваксинации срещу опасни и смъртоносни болести, като едра шарка, полиомелит и туберкулоза, тъй като в момента са рядко срещани. Те обаче (съзнателно или не) забравят, че тези болести са редки, тъй като всички или поне повечето от хората са ваксинирани срещу тях. „Е“ да, ама в (някоя слабообразована страна) повечето не са ваксинирани и не са измрели“, биха възразили противниците на ваксинирането. Хората не са измрели, но въпреки това починалите от тези болести са повече и те не са включени в статистиката. По същия начин в ранните фази на епидемията от COVID-19 държавите в Азия



Фигура 2. Шлем Броди

и Европа затвориха училищата и много малко деца се заразиха. Това накара дори някои статистици да заключат, че децата много трудно се заразяват и не пренасят болестта – нещо, което се опроверга по-късно с изследвания и реални данни за заразените деца в САЩ и Южна Корея.

Друг пример са хората, които твърдейки, че цигарите или алкохолът не са толкова вредни, казват: „*Леля ми/баба ми/чичо ми пушеха/пиха, но доживяха до 88-годишна възраст*“. Доказано е, че хората, които, водейки нездравословен начин на живот, достигат до преклонна възраст, са малко и повечето умират рано. Въпреки това случаите на починали рано пушачи и/или алкохолици не се разглеждат, а малцинството случаи се представя като мнозинство. Друг пример дава Насим Талеб в книгата си „Надхитрени от случайността“ (Taleb, 2005). В даден инвестиционен фонд пристига трейдър, който има пет последователни успешни години и печели стотици милиони за компанията, но на шестата година се проваля и губи повече, отколкото е спечелил до момента, с което тя остава на нетна загуба от него. Как е възможно това? Насим Талеб дава прост пример.

Нека предположим, че трейдърът успява само благодарение на късмета, с вероятност за успешна година от 50%. Вероятността за пет поредни години е произведение $\frac{1}{2}^5$ или $\frac{1}{32} \approx 3\%$. Тя изглежда малка, но всичко зависи от извадката. На борсите работят десетки хиляди души. Ако хиляда от тях разчитат изцяло на

шанса, средно $\frac{1000}{32} = 31.25$, ще успяват пет поредни години. Шефовете виждат един-единствен човек, но не и цялата извадка и си мислят, че е много малко вероятно един мошеник да ги лъже толкова години поред (Kahneman, 2013).

Магията на Валд

Интересен е начинът, по който Валд пресмята броя на свалените и оцелелите самолети. Той използва следния подход, който тук даваме с конкретен пример, взет от (Mangel and Samaniego, 1984), предадено от (Bil Casselman, 2016). Ползвайки група от 400 самолета, той обозначил всичките 400 машини с N . От 400 самолета се завърнали 380. Те били обозначени с S и представлявали 95% от общия брой самолети. Свалените самолети L са 20 на брой, съставляват 5% от общия брой самолети L . С S_0 се отбелязват завърналите се без дупки от снаряди самолети. Те са 320 и са 80 процента от цялата група. С една дупка са 32 самолета (8%) от групата. С две дупки са 20 самолета, или 5 процента от самолетите. С 4 и 5 дупки са по два самолета, съответно 0,5% (табл. 1).

Таблица 1. Числен пример за работата на Валд

	Брой	Съотношение
Самолети в мисията	400	1
Завърнали се	$N = 380$	0.95
Свалени	$L = 20$	0.05
Завърнали се без попадения	$S_0 = 320$	$s_0 = 0.8$
С едно попадение	$S_1 = 32$	$s_1 = 0.08$
С две попадения	$S_2 = 20$	$s_2 = 0.05$
С три попадения	$S_3 = 4$	$s_3 = 0.01$
С четири попадения	$S_4 = 2$	$s_4 = 0.005$
С пет попадения	$S_5 = 2$	$s_5 = 0.005$

Валд приема, че всички свалени самолети са свалени от вражески огън, а не от технически повреди. С L_i се отбелязват свалените с даден брой попадения самолети, означен с i . L_1 са самолетите, свалени с едно попадение, L_2 – с 2, L_3 – с 3, и т.н. Общият брой свалени или загубени самолети (lost) (1) е сборът от всички категории на загубени самолети – с едно попадение, с две, с три

и т.н. Броят на всички загубени самолети е разликата между първоначалния (общ) брой N и крайния брой на завърналите се самолети S :

$$L = \sum_i L_i = N - S \quad (1)$$

Също както загубените самолети, така и общият брой и броят на завърналите се могат да се разделят на категории по броя попадения, който имат. Във всяка категория (например самолети с 5 попадения) броят на загубените L_i и броят на завърналите се самолети S_i заедно дават общия брой самолети в тази категория (2):

$$L_i + S_i = N_i \quad (2)$$

Разделянето на самолетите по категории може да се направи по много начини. Ако искаме например да разделим самолетите на две групи – една група с брой попадения h под някаква стойност $h = i$ $N_{h < i}$ и втора група самолети с поне толкова попадения $h \geq i$ $N_{h \geq i}$. Групата на самолетите с поне i попадения е обединение на групите от самолети с 0 попадения, с 1 попадение, с 2 попадения с $i - 1$ попадения. В нашия пример можем да групираме самолетите на две групи по много начини, един от които е: самолети с поне 1 попадение $N_{h \geq 1}$ и самолети без попадения $N_{h < 1}$. Броят на самолетите с поне едно попадение е сбор от този на самолетите с 1, 2, 3, 4 и 5 попадения, а броят на самолетите с поне i попадения е сбор на броя самолети с $i + 1, i + 2, i + 3$ и т.н. попадения (3):

$$N_{h \geq i} = \sum_{j \geq i} N_j \quad (3)$$

Сборът от двете групи ще даде общия брой (4), както и да разделим на две групи различните категории самолети:

$$N_{h < i} + N_{h \geq i} = N \quad (4)$$

Като знаем броя на завърналите се с i попадения самолети S_i и броя на падналите с i попадения L_i , разбираме общия брой на всички самолети (независимо дали изгубени, или не) в тази категория:

$$S_i + L_i = N_i \quad (5)$$

Целта на Валд е да разбере колко самолета са паднали с 1 попадение, колко с 2 и т.н., т.е. да намери всяко L_i или поне да *оцени* този брой – да го постави в някакъв интервал, в който той попада с висока вероятност – доверителен интервал. Това е трудна, ако не и невъзможна задача с директен подход, тъй като това са ненаблюдаеми и непознаваеми количества. Това, което може да

се направи, е да се намерят *оценки* за вероятностите за падане на самолети на i -тото попадение, при условие че не са паднали по-рано. С p_i се обозначава условната вероятност на самолета да падне на i -тото попадение, след като преди това е оцелял $i - 1$ попадения. Например:

p_1 е вероятност самолетът да е паднал при първото попадение, p_2 – при второто попадение, след като е оцелял на първото, p_3 – при третото попадение, след като е оцелял на второто, и т.н.

Валд тук прави едно опростяване – приема, че всеки самолет с даден брой попадения има равен шанс да падне, така вероятността става просто съотношение на благоприятни събития към общ брой възможни събития (6):

$$p_i = \frac{L_i}{N_{h \geq i}} \quad (6)$$

Числителят на дробта е броят загубени самолети с i попадения, а знаменателят ѝ е общият брой самолети (загубени и върнати) с поне i на брой попадения. Защо в знаменател е $N_{h \geq i}$, а не само N_i ? Защото всички самолети, които биха могли да паднат на i -тото попадение, имат i попадения. Самолетите с 3 попадения са тези с 3 попадения, с 4 попадения, с 5 попадения – всяко число съдържа *предшното*. Получилите 6 попадения *са имали възможността* да паднат на 3-тото. Оттам всички самолети с поне i на брой попадения формират множеството на всички благоприятни събития, към които трябва да съотнесем реално случилите се. По същия начин пресмятаме вероятността, когато хвърлим зар, да се падне нечетно число – 3 благоприятни събития (1, 3 или 5) от 6 възможни – $p_{n=2k+1} = \frac{3}{6}$.

Изразът (6) е началната точка на нашето изследване. За момента в него нямаме известни променливи. Трябва да го преобразуваме, за да намалим неизвестните и да обвържем намирането на дадено L_i с известни към момента променливи. Това означава по възможност L_i да зависи от L_{i-1} , с което да получим рекурсия, с която можем да изчислим броя на загубените самолети например с 5 попадения от броя на загубените 4, а тях от броя на загубените с 3 и така до началото – загубените самолети без попадения – брой, който може да се оцени на база на техническите им характеристики, без да имаме нужда да знаем как са били обстрелвани, и да зависим от непознаваеми величини. Затова преобразуваме:

$$p_i = \frac{L_i}{N_{h \geq i}} = \frac{L_i}{\sum_{j \geq i} N_j} \quad (7a)$$

$$L_i = p_i \left(\sum_{j \geq i} N_j \right) = p_i \left(N - \sum_{j < i} N_j \right) \quad (7b)$$

В (7б) обърнахме индексите – вместо да търсим сбора на всички самолети с поне i попадения, използваме разликата между общия брой самолети и тези с по-малко от i , което е същото нещо. Това преобразуване ни позволява обаче да направим зависимост между загубените самолети с даден брой попадения и загубените самолети с по-малко попадения:

$$L_i = p_i \left(N - \sum_{j < i} N_j \right) = p_i \left(N - \sum_{j < i} S_j - \sum_{j < i} L_j \right) \quad (7в)$$

В (7в) вече има само две група неизвестни (върналите се самолети по категории и общият брой самолети са дадени). Една от тях са падналите самолети от различни категории (различен брой попадения), при които има рекурсивна зависимост – за да сметнем например L_5 , трябва да знаем $L_0, L_1 \dots L_4$. Така можем да се върнем една стъпка назад и сега, за да сметнем L_4 , трябва да знаем $L_0, L_1 \dots L_3$. Накрая стигаме до L_1 , който зависи само от L_0 . Валд прави допускане, че самолетите падат само от вражески огън, с което опростява изчислението $L_0 = 0$. Това ни позволява да изчислим всички L_i последователно.

$$\begin{aligned} L_1 &= p_1(N - S_0) \\ L_2 &= p_2(N - S_0 - S_1 - L_1) \\ L_3 &= p_3(N - S_0 - S_1 - S_2 - L_1 - L_2) \dots \end{aligned} \quad (7г)$$

Остава само един проблем – не знаем вероятностите p_i и няма как да ги узнаем директно. Абрахам Валд е работил, за да получи оценки за техните стойности. В задачи от теория на вероятностите, когато е трудно и сложно да се оцени вероятността да се случи едно събитие, превключваме на вероятността то да НЕ се случи, тъй като те заедно дават 1:

$$P(A) + P(\bar{A}) = 1 \quad (8)$$

Вероятността да се случи едно събитие или да не се случи, е 1, тъй като няма трета възможност. Това е законът за изключеното трето *tertium non datur* – Садам Хюсеин или е жив, или е мъртъв. В известен смисъл двете вероятности (на събитието и неговото несбъждане) са обвързани – когато едната се смята много сложно, тъй като е комбинация от много други събития, другата се смята много просто. Затова образуваме вероятността самолетът да НЕ падне на i -тото попадение, т.е. да оцелее, до i включително (9а):

$$q_i = 1 - p_i \quad (9а)$$

$$q_i = 1 - \frac{L_i}{N_{h \geq i}} = \frac{N_{h \geq i} - L_i}{N_{h \geq i}} \quad (9б)$$

$$q_i N_{h \geq i} = N_{h \geq i} - L_i \quad (9в)$$

Тук търсим отново рекурсивна зависимост, но за общия брой самолети с повече от i попадения, която да не включва изгубените самолети, за да имаме само един тип неизвестни. Затова трябва да преобразуваме разликата в (9в), така че да участват върналите се самолети вместо изгубените. Как може да стане това? Общият брой самолети над дадено ниво на попадения $N_{h \geq i}$ е сбор на броя загубени самолети над това ниво и броя върнали се (9г):

$$N_{h \geq i} = S_{h \geq i} + L_{h \geq i} \quad (9г)$$

Броят завърнали се самолети с поне i попадения $S_{h \geq i}$ е сбор от различни категории самолети също както общият брой (3). Можем да използваме това, за да направим рекурсивна връзка между $N_{h \geq i}$ и $N_{h \geq i+1}$, в която участват вероятностите q_i , която да ни позволи да ги изчислим всяка от предишната. За да го направим, трябва да отделим самолетите с най-малкия брой попадения в (9г):

$$N_{h \geq i} = S_{h \geq i} + L_{h \geq i} = S_i + L_i + S_{h \geq i+1} + L_{h \geq i+1} = S_i + L_i + N_{h \geq i+1} \quad (9д)$$

Това ни позволява да преобразуваме (9в) в удобна за нас форма:

$$q_i N_{h \geq i} = N_{h \geq i} - L_i = S_i + L_i + N_{h \geq i+1} - L_i = S_i + N_{h \geq i+1} \quad (9е)$$

В уравнение (9е) имаме рекурсия също както в (7в) и отново две неизвестни. Нужни са ни допълнителни допускания, за да можем да решим задачата. Едно от тези допускания, направени от Валд, е, че има горна граница на броя попадения, които един самолет може да получи – в нашия пример това са 5 удара. Продължителността на престрелката, скоростта на летене и други фактори поставят тази граница както за падналите, така и за върналите се самолети. Опростяването в случая е, че тази граница n , за която $N_{h > n} = 0$, е една и съща за падналите и върналите се самолети, което противоречи на здравия разум, който ни казва, че падналите самолети би трябвало да имат средно повече попадения от завърналите се. Въпреки всичко това и други неинтуитивни допускания са позволили на Валд да изчисли добри оценки за вероятностите q_i . Ако той бе допуснал разлика в тази граница, това би въвело още едно непознаваемо количество, свързано с липсващите самолети, нуждаещо се от оценяване, с което задачата би станала много по-трудна. С тази граница се получава лесна рекурсия, тъй като когато $i = n$, (9е) се опростява до:

$$q_n N_{h \geq n} = S_n + N_{h \geq n+1} = S_n \quad (9ж)$$

Самото (9е) може да се преобразува така, че да изрази рекурентната връзка, която търсим:

$$N_{h \geq i} = \frac{S_i}{q_i} + \frac{N_{h \geq i+1}}{q_i} \quad (93)$$

Тази формула, макар и рекурсивна, отново не ни дава директен начин да изчислим вероятностите q_i . От тези количества $N_{h \geq i}$ (общ брой самолети с поне i попадения) в почти всички участват паднали самолети с определен брой попадения, които са неизвестни величини. Единственото количество, което ни е известно, е общият брой самолети, събиращ всички категории

$$N = N_{h \geq 0} \quad (94)$$

Силата на математиката се корени в абстрактното мислене и способността да се манипулират символи – когато не можем да използваме рекурсията, за да изчислим директно, можем да я използваме, за да изведем изчислим израз. За целта тръгваме от началото (9ж) и заместваме последователно в рекурсията (9з) с изрази (символни низове):

$$\begin{aligned} N_{h \geq n} &= \frac{S_n}{q_n} \\ N_{h \geq n-1} &= \frac{S_{n-1}}{q_{n-1}} + \frac{N_{h \geq n}}{q_{n-1}} = \frac{S_{n-1}}{q_{n-1}} + \frac{S_n}{q_{n-1}q_n} \\ N_{h \geq n-2} &= \frac{S_{n-2}}{q_{n-2}} + \frac{N_{h \geq n-1}}{q_{n-2}} = \frac{S_{n-2}}{q_{n-2}} + \frac{1}{q_{n-2}} \left(\frac{S_{n-1}}{q_{n-1}} + \frac{S_n}{q_{n-1}q_n} \right) = \frac{S_{n-2}}{q_{n-2}} + \frac{S_{n-1}}{q_{n-2}q_{n-1}} + \frac{S_n}{q_{n-2}q_{n-1}q_n} \\ N_{h \geq 0} = N &= S_0 + \frac{S_1}{q_1} + \frac{S_2}{q_1q_2} + \dots + \frac{S_n}{q_1q_2\dots q_n} \end{aligned} \quad (10a)$$

Уравнение (10) е формулата на Валд, с която той изчислява вероятностите за оцеляване. То може да се нормира, като вместо брой се използват съотношения:

$$\begin{aligned} N - S_0 &= \frac{S_1}{q_1} + \frac{S_2}{q_1q_2} + \dots + \frac{S_n}{q_1q_2\dots q_n} \\ \frac{N-S_0}{N} &= \frac{s_1}{q_1} + \frac{s_2}{q_1q_2} + \dots + \frac{s_n}{q_1q_2\dots q_n}, \quad S_i = \frac{S_i}{N} \\ 1 - s_0 &= \frac{s_1}{q_1} + \frac{s_2}{q_1q_2} + \dots + \frac{s_n}{q_1q_2\dots q_n} \end{aligned} \quad (106)$$

Тук имаме едно уравнение с n променливи, което не може да се реши, ако те са независими. Валд прави второ допускане, което е неинтуитивно. Здравият разум ни казва, че всяко попадение върху самолета намалява вероятността той да оцелее, т.е.

$$q_1 \geq q_2 \geq q_n \quad (11)$$

Според (Casselman, 2016) Валд приема, че вероятността за оцеляване не намалява с увеличаване на броя попадения, което е една възможна интерпретация на заместването на всички тези вероятности с една-единствена (12):

$$q_1 q_2 \dots q_i = q^i \quad (12)$$

Това е т.нар. средно геометрично, дошло от „Елементи“ на Евклид. Заместването на тези произведения с една усреднена вероятност на дадена степен може да се интерпретира и по друг, по-разумен начин – вероятностите са близки една до друга – един допълнителен удар не променя скокообразно вероятността за оцеляване. Средно геометричното „наказва“ големите отклонения на данните, което филтрира систематични грешки при изчисляване. То е най-подходящо при пресмятане средна стойност на произведения на съотношения. С това допускане уравнението се опростява до такова с една неизвестна и може да се реши числено или аналитично:

$$1 - s_0 = \frac{s_1}{q} + \frac{s_2}{q^2} + \dots + \frac{s_n}{q^n} \quad (106)$$

В конкретния пример вероятността за оцеляване на първото попадение е $q = 0.85$, а за оцеляване на 5-ото попадение е $q^5 = 0.444$. Грешките от изчисление тук се умножават, поради което по-точните оценки са за вероятностите за оцеляване с по-малко попадения, най-точна е при $q = q_1$. Оттук пресмятането на вероятностите за сваляне на самолети на i -тото попадение, при условие че са оцелели на $i - 1$, е тривиално (9а), а с тях можем да оценим броя на загубените самолети по брой попадения (7в).

Заклучение

Работата на Абрахам Валд по шансовете за оцеляване на самолети при престрелки не се изчерпва с това уравнение, то е началото на дълго изследване, резюмирано в (Mangel & Samaniego, 1984), състоящо се в получаване на граници, в които вероятностите за сваляне на самолетите варират. Границите са необходими, тъй като има няколко ключови допускания, за да се получи уравнение за непознаваеми величини (106). Изчисляването на стойности за такива величини носи много по-малко информация, отколкото границите, в които те могат да варират. Това е начална фаза на едно изследване. Всяко истинско научно изследване, свързано с данни, започва от прости ситуации с множество допускания. При успех допусканията биват разхлабвани едно по едно и в последователни стъпки уравненията се усложняват и стават по-точни и приложими. Това е целият път на физиката – от уравненията на Нютон за точки с маса, през неговото доказателство за привличането на сфери с маса

(Tomov et.al, 2019). Валд действително се е занимавал и с връзката между местоположението на пораженията върху самолетите и вероятностите за тяхното завръщане, но това е само малка част от неговия съществен принос. В (табл.2) са показани резултати от неговия анализ за вероятността за оцеляване от едно попадение според мястото на поражение.

Таблица 2. Вероятност за оцеляване при едно попадение

Място на попадението	Стойност
Целият самолет	0.851
Двигател	0.588
Фюзелаж	0.940
Горивна система	0.973
Други	0.939

Изводите за конструкторите са ясни дори само от това изчисление – разликата между вероятността за оцеляване при поражение на двигателя и всички други части на самолета е съществена, а те са близки помежду си. Това е достатъчна информация за взимане на решение при управление на риска (Taleb, 2020). Известната история за Валд е по-скоро вярна, но е много малка част от направеното от него и звучи много по-просто, отколкото е действителното емпирично изследване, както всяка друга фолклорна история, свързана с учени и техните изследвания.

БЕЛЕЖКИ

1. източник: https://en.wikipedia.org/wiki/Survivorship_bias#/media/File:Survivorship-bias.png

ИЗТОЧНИЦИ

<http://www.ams.org/publicoutreach/feature-column/fc-2016-06>
https://www.researchgate.net/publication/254286514_Abraham_Wald's_Work_on_Aircraft_Survivability
<https://nauka.bg/bg/nauka-matematika/>

REFERENCES

Bill Casselman, The Legend of Abraham Wald, 2016, American Mathematical Society,
D. Kahneman, Thinking, Fast and Slow, Farrar, Straus and Giroux; 1st edition (April 2, 2013).

- M. Mangel, F.J. Samaniego, Abraham Wald's Work on Aircraft Survivability, 1984. Journal of The American Statistical Association – J AMER STATIST ASSN. 79. 259-267. 10.1080/01621459.1984.10478038.
- N.N.Taleb, Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets (Incerto), Random House Trade Paperbacks; Updated edition (August 23, 2005).
- N.N. Taleb, Statistical consequences of fat tails, Real World Preasymptotics, Epistemology, and Applications (Technical Incerto), STEM Academic Press; Illustrated edition (June 30, 2020), ISBN-10: 1544508050.
- L. Tomov et. al, Istoriya na matematikata, Balgarska nauka, November 2019, pp 183, ISSN: 1314-1031.

ABRAHAM WALD AND STATISTICAL RESEARCH ON AIRCRAFT ARMOR

Abstract. This work investigates the actual and the legendary scientific research of Abraham Wald on the survival probabilities of planes with different number of hits from enemy fire. We give new explanation to the deduction of his basic equation. We explain his basic assumptions and their role in the process of derivation, which we display in more detail. We investigate the role of intuition in this statistical study and how ignoring it allows to formulate assumptions and derive equation about missing data, which is unknowable in the strict sense of the term. We discuss the roles of estimations for unknowable or unobservable variables and highlight the methodology of sequential relaxations of estimations for recursive approximation of real-world data by successive theoretical models.

Keywords: history; Wald; heuristics; statistics; probability; epistemology

✉ **Prof. Dr. Latchezar P. Tomov**

ORCID ID: 0000-0003-1902-6473

New Bulgarian University

21, Montevideo Blvd.

1618 Sofia, Bulgaria

E-mail: lptomov@nbu.bg

✉ **Mr. Georgi Doychinov**

National School for Ancient Languages and Cultures

“Saint Constantine-Cyril the Philosopher”

16, Baba St.

Sofia, Bulgaria

E-mail: g.v.doychinov@gmail.com