$$a^2 + b^2 = c^2$$

ALGUNAS REFLEXIONES SOBRE LA DISTRIBUCIÓN BINOMIAL NEGATIVA II COMO INSTRUMENTO DE MEDICIÓN

(Un Análisis Teórico y Aplicado)

ISADORE NABI

ÍNDICE

I. PRÓLOGO

La presente investigación es un esfuerzo por abordar de una forma dialéctica el estudio de las Probabilidades y la Teoría Estadística, específicamente el estudio de la distribución binomial negativa 2 (NB2). Esto implica estudiar las distribuciones de probabilidad en sus conexiones mediatas e inmediatas con otras familias de distribuciones de probabilidad en términos históricos, lógico-formales, en términos de la multiplicidad de marcos teóricos científicos que ofrezcan las ciencias a la Teoría Estadística, así como también la respectiva multiplicidad de investigaciones aplicadas derivadas de tales marcos teóricos, todo ello interrelacionado mediante la lógica dialéctica. Esta lógica, aunque aquí no es el lugar en el que se especificará su significado teórico, desea el autor sí pueda ser comprendida intuitivamente en su dimensión aplicada, en tanto esta investigación constituye un modesto esfuerzo por mostrar esta última dimensión de la dialéctica. Un esfuerzo que el lector deberá juzgar como considere pertinente con base a un estudio riguroso del mismo, por lo cual de antemano el autor está profundamente agradecido porque comprende las implicaciones de ello.

Por otro lado, la investigación está diseñada para que, en general, un lector que no desee profundizar en aspectos que son relativamente tangenciales a la investigación pueda desenvolverse en ella sin necesidad de recurrir a las notas al pie. Sin embargo, se recomienda que se revisen y con base en ello se valore si estudiarlas o no, porque son en ellas donde, en general, reside la crisálida de las intenciones filosóficas de esta investigación.

La investigación posee dos ejes fundamentales. El primero, que es al que aquí se le ha asignado mayor relevancia, es el estudio de la NB2 como parte de un cuerpo teórico más amplio, específicamente el de los modelos jerárquicos y las mixturas de distribución que más adelante se estudiarán. El segundo de estos ejes fundamentales es el estudio de la NB2 como distribución de probabilidad individualmente estudiada a nivel teórico y como distribución de probabilidad aplicada a la resolución de problemas prácticos, para lo cual se resuelve un caso aplicado manualmente, mediante una calculadora y mediante el paquete estadístico R Studio.

En el transcurso de la investigación, se va buscando una unificación intuitiva y lógica entre la aparición histórica y la construcción matemática de los instrumentos de medición utilizados para modelar el comportamiento de fenómenos naturales o sociales en condiciones de incertidumbre, es decir, de las distribuciones de probabilidad. El autor no se limitó en esfuerzos por fundamentar con la suficiente cantidad de literatura filosófica y científica sus puntos de vista, independientemente de que algunos puedan resultar (o no) polémicos, así como tampoco en que no existan malos entendidos ni buenos entendidos, sino únicamente entendidos o no entendidos. Por supuesto, la realidad no es en general dicotómica y tales esfuerzos son un esfuerzo de minimización.

Se aborda en diferentes secciones lo relacionado con el Teorema de Bayes y el Bayesianismo Objetivo, con la finalidad de unificarlo dentro del cuerpo lógico de la Dialéctica Materialista; sobre ello se hace particular hincapié en los anexos de esta investigación. Así, el espíritu que anima esta investigación es contribuir a los esfuerzos de JBS Haldane, Richard Levins, Andrei Kolmogórov, Anatoliy Skorokhod, Richard Lewontin, Stephen Jay Gould, Lev Landau, M. M. Rosental, P. F. Iudin, Nikolái Bujarin, Roman Rosdolsky y otros pensadores por la consolidación del Marxismo como Filosofía de las Ciencias[1].

---

[1] "Soviet scholarship has produced an enormous inventory of publica- tions in the history of science. Only a small part of this output has been cast within the framework of carefully and consistently elaborated Marxist theory. The search for such a theory has been an exceedingly slow- and unending-process, kept off a steady course by a rapid succession of methodological turns, thematic shifts and ideological mutations." (Vucinich, 1982). Esto se puede decir para todo el Marxismo actualmente, lo cual es debido en parte a que el núcleo filosófico del Marxismo no está bien definido y si el lector consulta la historia sobre la obra de Marx (en particular de las condiciones de publicación del Tomo II y del Tomo III de *El Capital*, así como también de los *Grundrisse* y de *Teorías Sobre la Plusvalía*, y profundiza un poco en ello a nivel teórico, comprenderá lo que aquí se afirma.

Existen visiones que, aunque no explícitamente dialécticas, tienen la suficiente riqueza para incorporarse a una visión dialéctico-materialista de las Probabilidades y la Teoría Estadística. Un ejemplo contundente de ello se encuentra en (Feller, 1968, págs. 1-3). En la localización referida se señala que "Probability is a mathematical discipline with aims akin to those, for example, of geometry or analytical mechanics. In each field we must carefully distinguish three aspects of the theory: (a) the formal logical content, (b) the intuitive background, (c) the applications. The character, and the charm, of the whole structure cannot be appreciated without considering all three aspects in their proper relation.". Lo anterior brinda una idea al lector de la importancia filosófica del prólogo de la célebre obra de William Feller, sin embargo, debe hacerse una aclaración.

Según Feller los 3 aspectos anteriores deben estar lo suficientemente separados como para generar un divorcio entre sí, de tal forma que implica que deben tomarse en cuenta sin una unificación lógica y eso en términos de esta investigación resulta profundamente equivocado y se ha considerado necesario señalarlo; como se señalará más adelante, incluso científico que descubrió-inventó la técnica de Bootstrap ha demostrado que existen diversos casos (y no sólo él, sin embargo, no se profundiza en ello en esta investigación) en que el enfoque filosófico que se le da a las probabilidades define la interpretación y validez de un resultado. A excepción de esa salvedad, el prólogo de Feller proporciona insumos valiosos para comprender la riqueza filosófica, teórica y aplicada de la Estadística en todas sus manifestaciones. Todo ello con la finalidad de mostrar que el divorcio entre Filosofía de la Estadística, Estadística Matemática y Estadística Aplicada no es en lo absoluto aceptable y que no existe nada más práctico que la Filosofía, si esta es verdaderamente científica.

II. LA SERIE GEOMÉTRICA, EL TEOREMA BINOMIAL Y LA SERIE BINOMIAL NEGATIVA

II.I. Antecedentes Históricos, Intuición y Aplicaciones Modernas de la Serie Geométrica

La serie geométrica es una de las sumas infinitas más conocidas y de amplia utilización en diversas ramas de las ciencias naturales y sociales. Los fundamentos históricos que dan origen al nacimiento de esta serie poseen profundas raíces filosóficas y científico-aplicadas, específicamente en la filosofía y práctica científica de la antigua Grecia. Es en la Proposición 35 de la célebre obra *Elementos* de Euclides[2] donde aparecen las bases teóricas que después evolucionarían hasta devenir en la serie geométrica que conocemos en la actualidad. Sin embargo, la

[2] Véase (Joyce, 2014).

serie geométrica también está íntimamente relacionada con las paradojas del movimiento de Zenón y el nacimiento del Cálculo Diferencial e Integral, así como también del Análisis Matemático (que puede verse como una generalización del primero mediante el uso de desigualdades -con todo lo que ello implica a nivel filosófico e instrumental-), puesto que las disciplinas anteriormente expuestas son el establecimiento lógico-formal (lo que un matemático llamaría "formalización") de la respuesta de Aristóteles (padre de la Lógica-Formal, por cierto) a las paradojas de Zenón: Una suma infinita de elementos puede tener una existencia finita. Tanto el Cálculo como el Análisis muestran que Aristóteles tenía razón, pero no explican por qué, lo requiere un conocimiento de la estructura de los números reales que actualmente parece que no se posee[3] y reflexionar sobre ello escapa a los objetivos de esta investigación. En relación a este último vínculo histórico, fue Arquímedes quien utilizó esta serie (en términos de suma) para calcular un área comprendida entre una parábola y una línea recta; esta aplicación tenía como lógica orquestadora de fondo el célebre Método por Agotamiento de Arquímedes, que es el método de cálculo de áreas y volúmenes que evolucionaría después en lo que conocemos como Cálculo Diferencial e Integral.

La intuición detrás de la serie geométrica es ser una sucesión en el que cada término es un promedio geométrico de su sucesor y su antecesor, tal y como se muestra en la figura presentada a continuación:

Figura 1

[3] Prueba de ello es la Hipótesis del Continuo de Cantor que sigue sin demostrarse (o refutarse) desde su aparición a finales del Siglo XIX.

A very beautiful and illuminating geometrical illustration of this convergence is given in Carslaw's *Plane Trigonometry*. Take a line-segment OA of unit length, and draw OP of gradient x and AP of gradient 1, meeting at

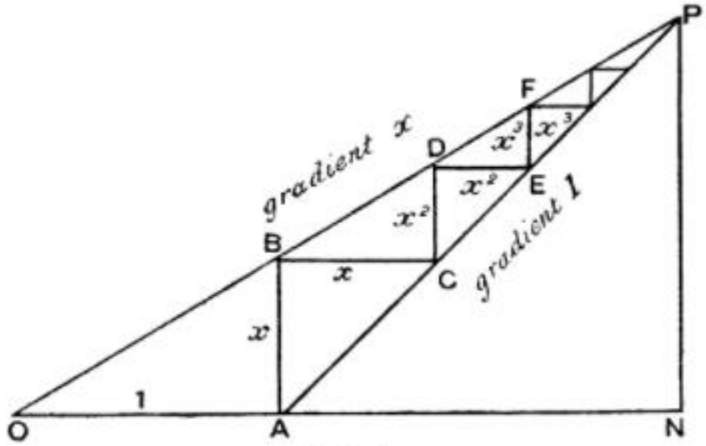


FIG. 4.

P. Then OA, AB, BC, CD, DE, EF, etc., drawn successively at right angles, have the values 1, x, x, x^2, x^2, x^3, etc. Draw PN perpendicular upon OA produced. Then ON is the limit of the sum $1+x+x^2+\dots$, and denoting this by s, $NP=s-1$. Clearly $s-1=xs$; whence $s=\dfrac{1}{1-x}=f$. The remainder $x^n f$ may also be obtained directly from the figure.

Fuente: (Stack Exchange, 2015).

Conforme las Matemáticas y sus aplicaciones evolucionaron en el tiempo, la serie geométrica fue objeto de distintas generalizaciones y fue encontrando cada vez más aplicabilidad en las prácticas científicas y técnicas de la humanidad. En la actualidad, la serie geométrica es una herramienta valiosa para resolver problemas científicos de distinta complejidad y tipo, entre los cuales están la estimación del valor del dinero en el tiempo en Finanzas, el cálculo de perímetros, áreas y volúmenes de objetos geométricos auto-similares en el estudio de Fractales, en la estimación de decimales periódicos en Métodos Numéricos, en el estudio del crecimiento de poblaciones en Biología, entre muchas otras.

II.II. Deducción del Teorema Binomial a partir de la Serie Geométrica

Como se sabe, la serie geométrica puede representarse de la siguiente forma:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \qquad (1)$$

Si se calcula la derivada de esta serie con respecto a x, se obtendría:

$$\frac{d}{dx}\left(\sum_{k=0}^{\infty} x^k\right) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x^2)}$$

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x^2)} \qquad (2)$$

Es un resultado ampliamente conocido que a medida se incrementa el orden de derivación se ve cada vez de forma más clara el patrón que resultante de esta derivación. Así, la n-ésima derivada de la serie geométrica toma la forma:

$$\frac{d^n}{dx}\left(\sum_{k=0}^{\infty} x^k\right) = \frac{d^n}{dx}\left(\frac{1}{1-x}\right) = \sum_{k=0}^{\infty} k(k-1)(k-2)\ldots(k-(n-2))x^{k-(n-1)}$$

$$= \frac{(n-1)!}{(1-x)^n} \qquad (3)$$

Nótese que, por un lado, la identidad (3) es la generalización de la identidad (2), mientras que por otro lado en la identidad (3) se conserva la condición $|x| < 1$, así como también $n \in \mathbb{N}$, lo que puede verificarse por inducción matemática.

Por otro lado, un resultado ampliamente conocido del Cálculo Fraccional es la identidad expuesta a continuación, la cual expresa el concepto de la n-ésima derivada de x^k:

$$\frac{d}{dx}(x^k) = kx^{k-1}, \qquad \frac{d^2}{dx}(x^k) = \frac{k!}{(k-2)!}x^{k-2}, \qquad \frac{d^3}{dx}(x^k) = \frac{k!}{(k-3)!}x^k, \ldots,$$

$$\frac{d}{dx}(x^k) = \frac{k!}{(k-n)!}x^k \quad (4)$$

Evidentemente la expresión anterior sólo puede tomar valores discretos puesto que se estableció la condición $n \in \mathbb{N}$ sobre el objeto matemático que se está trabajando. Sin embargo, es también evidente que en el análisis de los fenómenos naturales y sociales lo finito sólo tiene sentido si es puesto en unidad con lo infinito, en una unidad que expresa la subsunción formal y real de lo finito por lo infinito (de las partes por el todo[4]). Por tanto, vamos a asumir que la expresión (4) es válida también para $n \in \mathbb{R}$.

Por tanto, si se asume que la expresión (4) es válida para $n \in \mathbb{R}$, es posible ampliar el dominio de la serie geométrica a los reales de tal forma que:

$$\frac{d}{dx}(x^k) = kx^{k-1}, \qquad \frac{d^2}{dx}(x^k) = \frac{k!}{(k-2)!}x^{k-2}, \qquad \frac{d^3}{dx}(x^k) = \frac{k!}{(k-3)!}x^k$$

Por tanto, es válido asumir que[5]:

$$\frac{d^n}{dx}\left(\frac{1}{1-x}\right) = \frac{(n-1)!}{(1-x)^n}, \qquad n \in \mathbb{R} \qquad (5)$$

[4] Si hablar de una solución a la Hipótesis del Continuo es una tarea ardua y casi seguramente una misión imposible (hasta donde Gödel comentó y nadie ha refutado -o confirmado- desde otro punto de vista), unificar a Comte y a Marx es seguramente una misión imposible (aunque una grata fortuna, en este caso), sin embargo, es innegable que el principio de la supremacía del Todo sobre las partes de August Comte tiene su "isomorfismo filosófico" en la obra de Marx y de ello ni el Marxista más primitivo podría expresar objeción alguna. Esto prueba (en consonancia con la Hipótesis del Continuo en Matemáticas, la Hipótesis del Medio Continuo en Mecánica de Fluidos, del Axioma de Elección en Metamatemáticas, entre otros) que la continuidad como una envoltura del Todo sobre las partes (que está conformado con el Todo) es un elemento de trabajo analítico involuntariamente irrenunciable en cualquier filosofía casi seguramente científica (están las excepciones como el Positivismo) y este rasgo en el caso Marxista a nivel de la Economía Política ("Sólo el trabajo crea valor" -que nótese que ya incluye al capital, pero subsumido por el trabajo y por ello concebido como trabajo pretérito-, que es para dicha topología su longitud característica r) tiene sus orígenes en Hegel, en cuyo sistema la Hipótesis del Continuo tiene nombre y apellido: la Idea Absoluta, concepto que más familiarmente al menos una vez hemos llamado casi seguramente cualquiera de nosotros como Dios o algún isomorfismo cultural del mismo.

[5] Véase (Hilfer, 2010), específicamente las identidades 2.5 y 2.7 de las secciones 2.1.2 y 2.1.4 (tituladas **Euler** y **Liouville**) en las páginas 8 y 10 del documento electrónico, respectivamente para todas las referencias duales realizadas.

Lo anterior a su vez es equivalente a decir:

$$\sum_{k=0}^{\infty} k(k-1)(k-2)\ldots(k-(n-2))x^{k-(n-1)} = \frac{(n-1)!}{(1-x)^n}, \qquad n \in \mathbb{R} \qquad (6)$$

Nótese que (6) puede asumirse como verdadera, que es solamente el resultado de una manipulación algebraica y sustituciones realizadas en (5), las cuales son válidas por las propiedades lineales del operador diferencial y porque se definió que las identidades previas eran aplicables al dominio de los reales. La identidad anterior puede entenderse como una aproximación a la tasa de cambio infinitesimal a la que una serie geométrica varía desde su punto inicial hasta su punto final; por definición tal proceso evolutivo del fenómeno representado mediante el desarrollo de la serie geométrica (expresado este último matemáticamente como una expansión) se encuentra comprimido en el resultado $\frac{1}{1-x}$, que es el resultado del desarrollo de la serie geométrica hasta ∞.

De lo anterior se desprende que, multiplicando a ambos lados por $\frac{1}{(n-1)!}$ que:

$$\frac{1}{(n-1)!}\sum_{k=0}^{\infty} k(k-1)(k-2)\ldots(k-(n-2))x^{k-(n-1)} = \frac{1}{(1-x)^n}, \qquad n \in \mathbb{R} \quad (7)$$

Realizando otra manipulación algebraica se obtiene:

$$(1-x)^{-n} = \sum_{0}^{\infty} \frac{k(k-1)(k-2)\ldots(k-(n-2))x^{k-(n-1)}}{(n-1)!} x^{k-(n-1)}, \qquad n \in \mathbb{R} \quad (8)$$

Nótese que en la suma anterior todos términos que corresponden a los índices $k \leq (n-2)$ son iguales a cero debido a la estructura de $k(k-1)(k-2)\ldots(k-(n-2))$ que siempre involucra sustraer 2 unidades del término k-ésimo. Por tanto, el primer valor k que tendrá la característica de no-nulidad será $k=(n-1)$. Si se establece que el coeficiente de $x^{k-(n-1)}$ se denotará como C_k, es decir, $C_k = \frac{k(k-1)(k-2)\ldots(k-(n-2))x^{k-(n-1)}}{(n-1)!}$, entonces:

$$(1-x)^{-n} = \sum_{k=n-1}^{\infty} C_k\, x^{k-(n-1)}, \qquad para\; k = n-1, C_k = 1 \quad (9)$$

Lo anterior implica que el primer término de la serie es 1, lo que es similar al resultado conocido en la serie geométrica cuando no arranca en cero, sino en algún otro valor. De ello se desprende que cuando $k = n$ y $C_k = n$, entonces $C_k = \frac{n(n+1)}{2}$. Así, resulta intuitivo generalizar la lógica anterior para estimar C_k cuando el contador es igual a $k = n + r$:

$$C_k = \frac{n(n+1)(n+2)\dots(n+r)}{(r+1)!}, \qquad para\; r > 0$$

Si se sustituyen cada uno de los términos obtenidos en (8) o en (9), tales identidades se pueden re-expresar en los siguientes términos:

$$(1-x)^{-n} = 1 + nx + \frac{n(n+1)}{2}x^2 + \frac{n(n+1)(n+2)}{3!}x^3 + \cdots$$

Así, si se sustituyen x por $-x$ y n por $-n$, se obtiene:

$$(1+x)^n = 1 + (-n)(-x) + \frac{(-n)(-n+1)}{2}(-x)^2 + \cdots \qquad (10)$$

Puede observarse que para los valores impares siempre existe un signo negativo para x^n, por lo que es posible re-expresar el objeto matemático anterior. Por ejemplo, en el tercer término de (10) se tiene $(-n)(-n+1)(-n+2)$ y es posible re-expresar este componente de (10) como $-(n)(n-1)(n-2)$ y la forma negativa de x^n (es decir, $(-x)^n$) permite, gracias a su signo negativo, cancelar el signo negativo que queda fuera de los paréntesis en $-(n)(n-1)(n-2)$. Por otro lado, para términos pares, por ej0065mplo, para el caso de $(-n)(-n+1)$, es válido re-expresarlo bajo la forma $n(n-1)$ y como no existe para las potencias pares el signo negativo en $(-x)^n$ (porque $(-1)^{2k} = 1$), entonces finalmente se obtiene el resultado del Teorema Binomial:

$$(1+x)^n = 1 + nx + \frac{(n)(n-1)}{2!}x^2 + \frac{(n)(n-1)(n-2)}{3!}x^3 + \cdots \qquad (11)$$

II.III La Serie Binomial Negativa

Como se señala en (Weisstein, MathWorld, 2020), si la identidad (11) se generaliza re-expresándola en términos de una sustitución de 1 por a y de n por $-n$, se obtiene:

$$(x+a)^{-n} = \sum_{k=0}^{\infty}(-1)^k \binom{n+k-1}{k} x^k a^{-n-k}, \qquad para\ todo\ |x| < a \qquad (12)$$

III. LA DISTRIBUCIÓN BINOMIAL NEGATIVA COMO MARGINALIZACIÓN DE UN MODELO JERÁRQUICO DE TIPO MIXTURA DE PROBABILIDAD BINOMIAL-POISSON-EXPONENCIAL

III.I. Definición de Probabilidad

Aunque este no es el contexto óptimo para hablar a profundidad de la Filosofía de la Estadística, es ampliamente conocido que "Despite the fact that the concept of probability is such a common and natural part of our experience, no single scientific interpretation of the term probability is accepted by all statisticians, philisophers, and other authorities." (DeGroot & Schervish, 2012, pág. 2).

Sin embargo, a pesar de ello las probabilidades son computacionalmente válidas (tanto a nivel teórico como empírico) gracias al marxista soviético Andrei

Kolmogórov[6], que bajo una concepción filosófica objetiva de la realidad (que implica que el azar existe únicamente como recurso epistemológico) les proporcionó un fundamento topológico y axiomático para su estimación en la obra (Kolmogórov, 1956), específicamente estableció los axiomas para su cálculo de la página 2 a la 3, de la página 3 a la 5 estableció las relaciones de tales axiomas con los datos experimentales y en adelante su obra es eminentemente topológica y analítica.

En (DeGroot & Schervish, 2012, pág. 5) se plantea que "This theory is correct and can be usefully applied, regardless of which interpretation of probability is used in a particular problem.", sin embargo, esto es casi seguramente cierto. Cuando se estudia el fenómeno de manera lo suficientemente profunda la relevancia de las diferencias entre las distintas definiciones filosóficas de probabilidad cobra

---

[6] Esta investigación no versa sobre Historia de las Matemáticas, sin embargo, con conocimiento de causa de lo polémico que puede ser la afirmación anterior en una sociedad de clases capitalista, simplemente debe referirse el lector a (Eremenko, 2020). Ahí se encontrará que la visión filosófica de Kolmogórov era dialéctico-materialista. De hecho, su definición de Matemáticas es la misma que Engels en *Dialéctica de la Naturaleza*, añadiendo a ella la Lógica Matemática. Emerenko cita fundamentalmente tres documentos. El primero y el segundo son las versiones en ruso (http://www.mathnet.ru/links/8cdd5dd921cd8a51ba9423f541a3118c/ppi67.pdf) y en inglés (https://link.springer.com/article/10.1134/S0032946006040107) de la investigación de Kolmogórov cuyo título traducido al español es *Controversias Contemporáneas Sobre la Naturaleza de las Matemáticas*, respectivamente; mientras que la tercera es una investigación publicada un viernes 22 de enero de 2016 por la Universidad de Ulyanovsk en Rusia, cuya autoría corresponde a los profesores Baranets and Veryovkin, la cual puede encontrarse en el portal de la página web de la institución referida bajo la siguiente dirección: http://staff.ulsu.ru/baranetz/files/2011/06/baranec_verevkin_koncep_matemat_kolmogorova.pdf. Además, es bien conocido que Kolmogórov (por la investigación de él ya referida), aunque decantado intuicionista, rechazaba las visiones extremas del intuicionismo y del formalismo de Hilbert, que va de la mano con que en la Unión Soviética predominó de forma sumamente generalizada el Constructivismo Matemático, mucho después de que la cacería ideológica que hizo Stalin terminara, mucho después de la muerte del tirano incluso. Tanto así que hasta la fecha es la corriente filosófica que domina las Matemáticas en Rusia y sus otrora países satélites y de ello puede dar cuenta el lector si traduce las páginas de Wikipedia en ruso referentes a Filosofía de las Matemáticas, Matemáticas y Estadística (sean tópicos de Estadística Matemática o Estadística Aplicada), pero también existían fuertes raíces constructivistas antes de la cacería y mucho antes que el constructivismo se conociera como tal, y es que el lector debe recordar que la primera Escuela de Matemáticas en Rusia fue fundada por el ilustre matemático Leonhard Euler en lo que conformó la segunda etapa más productiva en el curso de sus investigaciones (la primera ocurrió en Alemania), particularmente en Moscú. Luego de ella sería nutrida por intelectuales y filósofos de la talla de Lobachevski y Chebyshov.

relevancia y cuenta de ello puede dar Bradley Efron, fundador de la técnica de re-muestreo conocida como Bootstrap y especialista en la utilización de métodos geométricos en la resolución de problemas de aplicación, en su investigación (Efron, 1978). Esta investigación puede sintetizarse de la misma forma en que el autor de la misma lo hace:

“Statistics, by definition, is uninterested in the special case. Averages are the meat of statisticians, where "average" here is understood in the wide sense of any summary statement about a large population of objects. "The average I.Q. of a college freshman is 109" is one such statement, as is "the probability of a fair coin falling heads is 1/2." The controversies dividing the statistical world revolve on the following basic point: just which averages are most relevant in drawing inferences from data? Frequentists, Bayesians, and Fisherians have produced fundamentally different answers to this question.

This article will proceed by a series of examples, rather than an axiomatic or historical exposition of the various points of view. The examples are artificially simple for the sake of humane presentation, but readers should be assured that real data are susceptible to the same disagreements. A counter-warning is also apt: these disagreements haven't crippled statistics, either theoretical or applied, and have as a matter of fact contributed to its vitality. Important recent developments, in particular the empirical Bayes methods mentioned in Section 8, have sprung directly from the tension between the Bayesian and frequentist viewpoint.” (Efron, 1978, pág. 232).

Establecido lo anterior, aquí se trabajará bajo una definición de probabilidades que se considera la noción básica de probabilidades sobre la que se debe trabajar con el fin de alcanzar una unificación filosófica de todas las Escuelas de la Filosofía de la Estadística sin ambigüedades. Por supuesto, la definición es en su estado actual aún insuficiente, por lo que deben hacerse algunas especificaciones al respecto.

En primer lugar, la definición que se presentará no incluye la Teoría de las Posibilidades (que es un desarrollo reciente en el análisis de la incertidumbre y no difundido de forma amplia), no permite tener completa claridad sobre en qué condiciones alguna de las corrientes aplica o no, así como tampoco especifica bajo qué condiciones puede variar la aplicabilidad de tales enfoques. Sin embargo, sí permite desterrar el concepto de subjetividad vista como una mera opinión, que tan difundido está entre los espíritus metafísicos que practican la Estadística, i.e., bayesianos subjetivos radicales. Con ello, permite sustituir en tal noción subjetiva el ver la probabilidad como una mera opinión por ver esta subjetividad como un criterio-experto basado en su conocimiento sobre el fenómeno estudiado y fenómenos similares (de la misma familia de fenómenos), la información disponible sobre dicho fenómeno y fenómenos similares (aunque sea poca o muy poca) y demás criterios subjetivos que tienen un fundamento eminentemente objetivo. Antes de pasar a exponer la definición, es necesario aclarar lo que aquí se quiere decir con que el azar es sólo un recurso epistemológico y para ello se retomará parcialmente la concepción de Poisson sobre el azar: "The law of large numbers is noted in events which are attributed to pure chance since we do not know their causes or because they are too complicated." (Poisson, 2013, pág. 16). A manera de comentario del traductor se señala que "Poincaré is known to have repeated that interpretation of randomness. For him, however, the main pattern of the action of chance was small causes leading to large consequences." (Poisson, 2013, pág. 29) y el lector debe recordar que Poincaré fue el precursor de la Teoría del Caos (antes del trabajo de Edward Lorenz en 1967), una teoría que trabaja en escenarios de una ausencia total de reglas (lo que en apariencia podría concebirse como antagónico al determinismo filosófico y científico), pero de espíritu completamente determinista, puesto que la lógica de sus fundamentos es isomórfica la lógica de Poisson sobre las probabilidades anteriormente expuesta. Finalmente, es conveniente mencionar que también tanto Jakob Bernoulli como Laplace (quien además es el fundador del determinismo científico) tenían una visión objetiva de la Estadística, hecho que parecen olvidar muchos bayesianos subjetivistas radicales.

Sobre Jakob Bernoulli puede verificarse la afirmación anterior en cuanto que "We here call this theorem Bernoulli's Theorem or Bernoulli's Fundamental Theorem. In the twentieth century, after strongers laws of large numbers had been proved, it also came to be known as the weak law of the large numbers (…) Bernoulli's fundamental theorem assumes that there is a fixed ratio of possible outcomes in a given situation, before going on to analyze what may occur when many outcomes in such a situation are observed (…) Not only *Ars Conjectandi* prove rigorously the first limit theorem in probability, it also founded the field of mathematical probability *conceptually*. For the first time it brought the epistemic concept of probability (*probabilitas*) into conjunction with the mathematics of games of chances. Before Bernoulli, the mathematics of games of chances had been developed by Pascal, Fermat, Huygens, and others largely without using the word (or concept of) "probability" (…)" (Bernoulli, 2006, págs. ix-xiii), lo cual corresponde a la nota de la traductora de la obra, Edith Dudley Sylla[7]; de lo Segundo da cuenta el hecho de que: "Todos los eventos, incluso aquellos que a causa de su insignificancia no parecen seguir las grandes leyes de la naturaleza, son el resultado de ella justo tan necesariamente como las revoluciones del sol. Ante la ignorancia de los lazos que unen tales eventos con todo el sistema del universo, se les ha hecho depender de causas finales o del azar, dependiendo de si ocurren y se repiten con regularidad o si aparecen sin respecto al orden; pero estas

[7] Dudley es profesora emérita de la Universidad Estatal de Carolina del Norte (véase (North Carolina State University, 2020) y fue profesora visitante de la Universidad de Radboud, Países Bajos, como puede verificarse en (Radboud Univeristy, 2011). Como se señala en la página de la segunda universidad, es célebre en los círculos académicos que investigan las Matemáticas Medievales y Modernas, la Historia de las Matemáticas (especialmente ratios y proporciones, el trabajo de Jacob Bernoulli -muy a la medida para el caso- y la transmisición de las matemáticas de los árabes -y sus recursos en Grecia e India- América Latina). Tan célebre es su trabajo que es una de los aspectos que se resaltan en la comercialización de la obra en Amazon, como puede verse en el siguiente enlace: https://www.amazon.com/-/es/Jacob-Bernoulli/dp/0801882354/ref=sr_1_2?dchild=1&qid=1601188550&refinements=p_27%3AEdith+Dudley+Sylla&s=books&sr=1-2.

causas imaginarias han retrocedido gradualmente ante los crecientes límites del conocimiento y han desaparecido por completo ante la sana filosofía, que sólo ve en ellas la expresión de nuestra ignorancia de las verdaderas causas." (Laplace, 2015, pág. 4).

Establecido todo lo anterior, finalmente es posible expresar que:

"La aplicación práctica de la Teoría de Probabilidades y de la Estadística Matemática se basa en el conocimiento de que el grado de indeterminación de la ocurrencia de sucesos aleatorios se puede determinar, para cada caso, de forma objetiva, mediante un número: la probabilidad. Para ello se parte, en correspondencia con la realidad objetiva, de que los fenómenos dependientes de la casualidad (entendida como recurso epistemológico en el sentido restringido de Poisson[8]), así como los procesos que transcurren de forma determinista, les son inherentes ciertas regularidades y de que la casualidad no significa ausencia total de reglas o caos. En este contexto se debe destacar que el concepto matemático de *probabilidad*, que define en forma objetiva y cuantitativa la probabilidad de un suceso aleatorio, se diferencia del concepto de lo *probable*, utilizado en el lenguaje común, que tiene generalmente fuertes caracteres subjetivos y con el cual muchas veces sólo se consideran proposiciones cualitativas. No obstante, se demuestra que las ideas subjetivas sobre la probabilidad de un suceso aleatorio se aproximan más y más a las relaciones objetivas que constituyen la esencia del concepto matemático de probabilidad, en la medida en que aumenta el arsenal de nuestras experiencias." (Maibaum, 1988, pág. 12)[9].

[8] "The law of large numbers is noted in events which are attributed to pure chance since we do not know their causes or because they are too complicated." (Poisson, 2013, pág. 16). Es restringido en el sentido de que la interpretación de Poisson involucra más que el contenido de la cita realizada, sin embargo, aquí se toma únicamente ello y se incorpora al cuerpo teórico planteado por Maibaum y lo mismo se hará respecto al concepto de probabilidad bayesiana objetiva planteados por Williamson, como se verá en los anexos de esta investigación.

[9] El contenido dentro de los paréntesis, así como estos, han sido añadidos por el autor de la presente investigación.

III.II Distribuciones Requeridas para este estudio de la NBII

III.I. Distribución de Bernoulli

Esta distribución de probabilidad es quizás la más elemental de todas en cuanto busca modelar el conjunto de eventos más simple posible: el que contiene únicamente $n = 1$ eventos. Esta distribución es la encargada de modelar la unidad experimental más elemental posible y, como se verá más adelante, es también parte de los resultados de las investigaciones de Jakob Bernoulli junto con la distribución binomial y otros elementos fundamentales de las Probabilidades y la Teoría Estadística. Precisamente la distribución binomial puede verse como una generalización inmediata de la distribución de Bernoulli ($n = 2$). Puede verse como el tránsito de un organismo unicelular a un organismo bicelular.

III.II. Distribución Binomial

Como se señala en (Maibaum, 1988, pág. 65), la distribución binomial se debe a Jakob Bernoulli (1654-1705), que fue uno de los primeros en trabajar la teoría de las probabilidades. Él como su igualmente célebre hermano Johann Bernoulli pertenecen a los más significativos discípulos de G.W. Leibniz (1664-1716). Jakob Bernoulli fue profesor desde 1687 hasta su fallecimiento en la Universidad de Basilea. En su obra *Ars Conjectandi* (publicada póstumamente en 1713), uno de los primeros libros sobre el Cálculo de probabilidades; este contiene proposiciones fundamentales, en particular, sobre la distribución binomial. Por eso se encuentra con frecuencia la distribución binomial bajo el nombre de distribución de Bernoulli (para múltiples ensayos independientes), y más aún la denominación del esquema de experimentos descrito como Esquema de Bernoulli, que consiste en n-ésimas repeticiones independientes del mismo experimento. Esto prueba las ligazones de esta distribución con los cimientos mismos de la Estadística Matemática y la Estadística Aplicada.

Así, la Distribución Binomial adopta la siguiente forma funcional[10]:

$$f(x|n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & para\ x = 0, 1, \dots, n \\ 0, & para\ cualquier\ otro\ valor\ de\ n \end{cases} \quad (13)$$

En esta distribución n es un entero positivo y oscila en el intervalo cerrado [0, 1], es decir, sólo existen dos posibles resultados en el experimento (éxito o fracaso, independientemente de cómo se definan). En (13), el coeficiente $\binom{n}{x}$ es el denominado *coeficiente binomial* y es igual a:

$$C_{n,x} = \binom{n}{x} = \frac{n!}{(n-x)!\,x!} \quad (14)$$

III.III. Distribución de Poisson

La distribución de Poisson aparece en un contexto histórico (Poisson nació en 1781) en que la teoría de las probabilidades estaba más desarrollada que cuando Jakob Bernoulli trabajó en la ahora conocida como Distribución de Bernoulli y en la Distribución Binomial (Jakob Bernoulli murió en 1705). En el trabajo de Poisson se puede ver ya una concepción más clara de las probabilidades que en Poisson (o menos poco clara, considerando que es un problema sin resolución hasta el día de hoy). Como es usual en el tipo de personalidades científicas que se han venido mencionando en esta investigación, las contribuciones de Siméon-Denis Poisson nutrieron teórica y aplicativamente diversos campos de las ciencias (integrales definidas, electromagnetismo y probabilidades) y precisamente fue en sus investigaciones sobre probabilidades en que aparece por primera vez el "Teorema

[10] Tomado de (DeGroot & Schervish, 2012, pág. 277).

Dorado” (así lo llamaba Jakob Bernoulli) llamado como la *Ley de los Grandes Números*.

Como se señala en (Encyclopaedia Brittanica, 2020), en la investigación que ya se citó en esta investigación, la cual versaba sobre los juicios criminales y civiles (lo que en el derecho contemporáneo llamaríamos “penales” y civiles), fue en donde nacen la distribución de probabilidad que lleva el nombre del ilustre científico, así como también en donde es llamada por primera vez la Ley de los Grandes Números (LGN) como tal. Cabe necesario resaltar que Poisson trabajó en una versión de la LGN para variables aleatorias independientes, con la misma distribución de probabilidad (fuese en masa o densidad) en que el valor promedio para muestras tendía a la media aritmética (es a lo que se refiere cuando se dice “a la media” en Estadística y también cuando se refiere al primer momento de probabilidad) a medida el tamaño de la muestra se incrementaba. Lo anterior es de importancia vital en esta investigación, puesto que en la obra de Poisson aquí referida se evidencia que originalmente la LGN era una mera aproximación a la distribución binomial y son estas nociones históricas lo que permiten que esta investigación sea “una sola pieza”, es decir, que nos permiten asumir a nuestra manera la hipótesis del continuo y saber que casi seguramente no habrán huecos lógicos ni expositivos relevantes (pensando en la convergencia a una media como un ponderado de los aspectos que como científicos consideramos deseables de una investigación). En los cimientos de la Estadística Matemática yacen la distribución de Bernoulli y la distribución binomial, junto con la distribución uniforme que apareció en la investigación titulada *Essay Towards Solving a Problem in the Doctrine of Chances* escrito por el reverendo Thomas Bayes en 1763 para el caso continuo y para el caso discreto en la obra del matemático calvinista Abraham De Moivre en 1718, como puede verificarse en (Kotz & van Dorp, 2004, pág. 3) y en (de Moivre, 1718, pág. 7), respectivamente.

Ahora bien, es fundamental comprender la necesidad histórica de la distribución de Poisson en cuanto descubrimiento de cierto tipo de fenómenos e invención de instrumentos para medir y teorizar a tales fenómenos. Como se señala en (Kim, Towards Data Science, 2019), investigadora senior en ingeniería en Microsoft, la distribución de Poisson surge como necesidad de modelar fenómenos naturales y sociales cuyo comportamiento contuviera más de dos posibles variantes dentro de la unidad de tiempo utilizada en el análisis de los mismos (que para la binomial es completamente dicotómico) y poder evitar el trabajo adicional (no necesariamente sencillo y/o simple) de particionar más y más la unidad de tiempo empleada, así como también permite estudiar los fenómenos en cuestión sin el requerimiento de conocer con antelación el número de ensayos[11].

Establecido lo anterior, es posible presentar al invitado estelar de esta sección. Al respecto, puede leerse en (DeGroot & Schervish, 2012, pág. 287) que "Many experiments consists of observing the occurrence times of random arrivals. Examples includes the arrivals of costumers for service, arrivals of calls at a switch-board, occurrences of floods and other natural and man-made disasters, and so forth. The family of Poisson distributions is used to model the number of such arrivals that occur in a fixed time period. Poisson distributions are also useful approximations to binomial distributions with very small success probabilities."

Los alcances de la distribución de Poisson son múltiples y eso sin considerar los denominados procesos de Poisson, sobre los cuales no se hablará por diversos motivos, aunque se recomienda al lector su estudio profundo. Sin embargo, puede comprenderse como una dinamización de la distribución de Poisson.

Sea $\lambda > 0$ la media de la distribución de la masa de probabilidad de un conjunto de observaciones. Los modelos cuya aleatoriedad epistemológica (en el sentido de

---

[11] En la fuente referida puede verse que "If use Binomial, you cannot calculate the success probability only with rate (i.e. 17ppl/week). You need "more info" (**n** & **p**) in order to use the binomial PMF", en donde PMF es la abreviatura en inglés de las Funciones de Masa de Probabilidad.

Bernoulli) es modelada con la distribución de Poisson, se asume que toman la siguiente forma, conocida como función de distribución de la masa de probabilidad[12]:

$$f(x|\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & para\ x = 0, 1, \dots , \\ 0, & para\ cualquier\ otro\ punto\ o\ valor\ de\ x \end{cases} \tag{15}$$

La esencia del nacimiento de la distribución de Poisson se encuentra en la necesidad de predecir eventos que ocurren en el futuro en un intervalo de tiempo fijo a una unidad temporal escogida. Por otro lado, en un proceso de Poisson ese parámetro, otrora la media, se convierte en una tasa de ocurrencia de eventos, como se explicará más adelante.

III.IV. Distribución Exponencial

Como debe ser norma en cualquier esfuerzo literario que busque alcanzar algún nivel mínimo respetable de cientificidad, antes de definir la estructura de la herramienta de medición utilizada se debe definir la estructura de los fenómenos o familia de fenómenos a medir con tal herramienta y, en este sentido, antes de definir matemáticamente un fenómeno o familia de los mismos debe también definirse en su intuición y lógica científica. Y la distribución exponencial también merece cierto detenimiento especial en el curso de esta investigación.

La distribución exponencial resuelve la necesidad de pronosticar la siguiente aparición de un fenómeno natural a determinada magnitud, con todo lo que ello puede abarcar. Para comprenderla en una dimensión relevante, se debe retomar la

---

[12] Este es el nombre canónico (en el sentido que describen sus orígenes filosóficos e históricos -con todo lo que ello a su vez implica-) de las conocidas no técnicamente como "distribuciones de probabilidad", a veces "distribuciones", a veces "masas" y otras "distribuciones de masa de probabilidad. Por su parte, están las funciones de distribución de la densidad de probabilidad para el caso continuo. El lector debe recordar que estos conceptos provienen de las Ciencias Físicas, específicamente de la Física Teórica, en donde la masa capta la intuición de ser la forma de la realidad que admite "huecos" (i.e., discontinuidades), mientras que la densidad es el concepto atinente a la realidad como un todo (conformada por materia, que es continua -de ahí la necesidad, y no sólo en la Física, de trabajar con la hipótesis del continuo descrita anteriormente en este trabajo-).

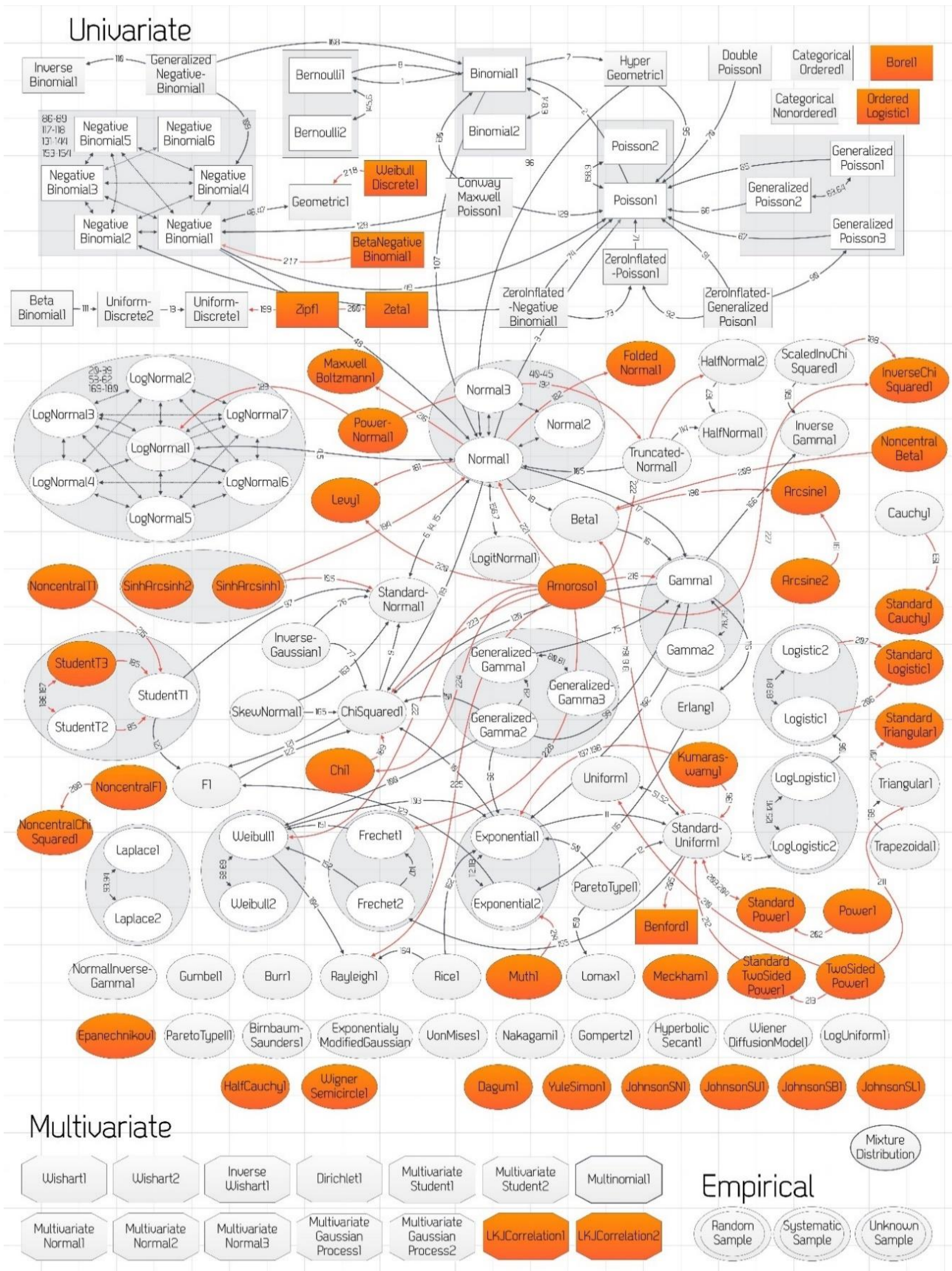
explicación de la sección anterior referente a la tasa de ocurrencia de eventos λ en un proceso de Poisson, que se definió intuitivamente como una dinamización de la distribución de Poisson, puesto que el tiempo entra a jugar un papel dinámico porque en lugar de ser el resultado de una medición es uno de los valores que componen el parámetro $\beta = \frac{1}{\lambda}$, en donde 1 representa la unidad temporal que se define cualitativamente y en donde λ es la media aritmética de una distribución de Poisson. En algunas fuentes bibliográficas, por ejemplo, en Wikipedia, puede encontrarse como λ, pero ahí la letra griega es concebida como la tasa o escala inversa, es decir, como el inverso multiplicativo del λ en la distribución de Poisson, por lo cual conviene mejor expresarlo en términos de la letra griega beta para evitar confusiones (aunque definido o conocido esto, no hay ambigüedades, puesto que los símbolos tienen el significado que los seres humanos les demos -el único cuidado que debe tenerse es que tal significado sea lo más acorde posible a la realidad estudiada y por eso nacen las ciencias-). Aquí se utilizará la notación presentada en (Wackerly, Mendelhall III, & Scheaffer, 2010), como se verá más adelante.

Como puede verificarse en la última fuente referida, la distribución exponencial también resuelve el problema del colapso del instrumento de medición cuando se mide un fenómeno cuya estructura se mide originalmente mediante la función de distribución gamma. Este es un ejemplo idóneo de cómo las longitudes características pueden colapsar en un sistema probabilístico como resultado de que el fenómeno estudiado sufre un salto cualitativo (como resultado a su vez de una acumulación cuantitativa de su desarrollo).

En palabras exactas de (Wackerly, Mendelhall III, & Scheaffer, 2010, págs. 185-186): "(…) α recibe a veces el nombre de *parámetro de forma* asociado con una distribución gamma. El parámetro β generalmente se llama *parámetro de escala* porque multiplicar una variable aleatoria con una distribución gamma por una constante positiva (y por tanto cambiando la escala en la que se hace la medición) produce una variable aleatoria que también tiene una distribución gamma con el mismo valor de α (parámetro de forma) pero con un valor alterado de β (…) En el caso especial cuando α es un entero, la función de distribución de una variable aleatoria con distribución gamma puede expresarse como una suma de ciertas probabilidades de Poisson (…) Si α no es un entero $0 < c < d < \infty$, es imposible dar una expresión de forma cerrada para $\int_c^d \left(\frac{(y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} \right) dy$ (…) Por esta razón cuando $\alpha = 1$ (una distribución exponencial), es imposible obtener áreas bajo la función de densidad gamma por integración directa. Valores tabulados para integrales como ésta se dan en *Tables of the Incomplete Gamma Function* (Pearson 1965).”

Lo anterior no sólo verifica la afirmación realizada previamente, sino también pone de manifiesto la relación existente entre la distribución de Poisson, el proceso de Poisson y la distribución exponencial, pero también entre la distribución exponencial y la gamma, y previamente se vio la relación entre la distribución de Bernoulli y la distribución binomial, así como de estas con la distribución de Poisson, es decir, lo anterior contribuye a verificar la interconexión teórica y lógica existente entre las distribuciones de probabilidad (de masa y de densidad). Si esta idea se generaliza con la ayuda de muchísima investigación y el uso intensivo de programas computacionales, es decir, si se desea establecer empíricamente las relaciones entre todas las distribuciones de probabilidad y sus re-parametrizaciones (para el caso univariante, de unas pocas multivariantes y de mixturas univariantes) es posible construir una gráfica como la que se presenta a continuación, en la que las distribuciones encerradas en color naranja representan las distribuciones que para el 13 de enero de 2017 representaban nuevas distribuciones de probabilidad descubiertas.

Figura 2



Fuente: (Swat, Grenon, & Wimalaratne, 2017, pág. 15).

Establecido lo anterior es posible plantear la función de distribución de densidad exponencial. Como puede verificarse en (Wackerly, Mendelhall III, & Scheaffer, 2010, pág. 212), se dice que una variable aleatoria X tiene una distribución exponencial con parámetro $\beta > 0$ si y solo si la función de densidad de Y es:

$$f(x) = \begin{cases} \dfrac{1}{\beta} e^{-x/\beta}, & para\ 0 \leq y \leq \infty \\ 0, & para\ cualquier\ otro\ punto \end{cases} \tag{16}$$

IV.VI. Modelos Jerárquicos y Mixturas de Probabilidad

IV.VI. I. Breves Conceptos Preliminares

IV.VI. I.I. Sobre el Ser En Sí y el Ser Para Sí

El sistema hegeliano sintetizó dialécticamente el estudio de la Naturaleza hecho por la Filosofía Griega y el estudio del Pensamiento hecho por el Idealismo Clásico Alemán. En ese sentido, Hegel denominó a la Naturaleza *Ser En Sí*, que puede entenderse bajo la expresión "Eso que está ahí" y al pensamiento como *Ser Para Sí*; el Ser En Sí puede entenderse en Hegel como la topología que determinada al objeto matemático estudiado, mientras que el Ser Para Sí puede entenderse como el objeto matemático en cuestión. La gramática misma delata que la Naturaleza y el individuo son concebidos como unidad, en que el individuo tiene una dimensión propia sólo en relación con un contexto más general a tal individuo. Desde el punto de vista de la lógica dialéctica-materialista también puede comprenderse como la cristalización de la noción de la existencia de una necesidad objetiva, en el estudio de los fenómenos naturales y sociales, de la consideración cualitativa y cuantitativa de un observador (que en Hegel es *Lo Particular*) dentro del contexto analítico (que en Hegel es *Lo Universal Abstracto*) en que se encuentra el fenómeno estudiado (su topología) y tras un proceso de investigación riguroso pueden obtenerse resultados cuya interpretación se corresponda con la realidad (que no es lo mismo que lo real -lo real es la forma física que toma la realidad, y si las formas fueran iguales a las esencias, la ciencia entera sobraría, dijo Marx en algún lado-) bajo criterios de verdad bien-definidos y que no pueden ser abordados de forma no rigurosa, para lo que esta investigación no es el escenario óptimo (eso sí, no pueden limitarse a ser lógico-formales ni basarse en ellos -el instrumento no puede dominar al investigador, así como en la mitología era concebido como fatalidad el hecho de que los demonios dominaran a sus conjuradores-).

A su vez, el proceso descrito anteriormente puede verse como el ampliamente conocido "Viaje del Retorno" en Epistemología Marxista, que es ir de *Lo Concreto* a *Lo Abstracto* y regresar a lo concreto dialécticamente, i.e., bajo la forma de *Lo Concreto Pensado*, que es el equivalente hegeliano de el *Ser En Sí-Para Sí*, la síntesis entre lo universal abstracto y lo particular, lo que genera el universal concreto[13].

---

[13] En palabras de Hegel: "El eliminar [Aufheben] y lo eliminado (esto es, lo ideal) representa uno de los conceptos más importantes de la filosofía, una determinación fundamental, que vuelve a presentarse absolutamente en todas partes, y cuyo significado tiene que comprenderse de manera determinada, y distinguirse especialmente de la nada. Lo que se elimina no se convierte por esto en la nada. La nada es lo inmediato; un eliminado, en cambio, es un mediato, es lo no existente, pero como resultado, salido de un ser. Tiene por lo tanto la determinación, de la cual procede todavía en sí (…) La palabra Aufheben [eliminar] tiene en el idioma [alemán] un doble sentido: significa tanto la idea de conservar, mantener, como, al mismo tiempo, la de hacer cesar, poner fin. El mismo conservar ya incluye en sí el aspecto negativo, en cuanto se saca algo de su inmediación y por lo tanto de una existencia abierta, a las acciones exteriores, a fin de mantenerlo. -De este modo lo que se ha eliminado es a la vez algo conservado, que ha perdido sólo su inmediación, pero que no por esto se halla anulado-. Las mencionadas dos determinaciones del Aufheben [eliminar] pueden ser aducidas lexicológicamente como dos significados de esta palabra. Pero debería resultar sorprendente a este respecto que un idioma haya llegado al punto de utilizar una sola y misma palabra para dos determinaciones opuestas. Para el pensamiento especulativo es una alegría el encontrar en un idioma palabras que tienen en sí mismas un sentido especulativo; y el idioma alemán posee muchas de tales palabras. El doble sentido de la palabra latina tollere (que se ha hecho famoso por la chanza de Cicerón: tollendum esse Octavium, Octavio debe ser levantado-eliminado) no llega tan lejos; la determinación afirmativa llega sólo hasta el levantar. Algo es eliminado sólo en cuanto ha llegado a ponerse en la unidad con su opuesto; en esta determinación, más exacta que, algo reflejado, puede con razón ser llamado un momento. El peso y la distancia respecto de un punto dado, se llaman en la palanca los momentos mecánicos de ella a causa de la identidad de su efecto, no obstante, todas las demás diferencias que hay entre algo real, como es un peso, y algo ideal, como la pura determinación espacial, es decir la línea. Véase Enciclopedia de las ciencias filosóficas 3ª edición, § 261, nota 9. Más a menudo todavía se nos va a imponer la observación de que el lenguaje técnico de la filosofía emplea para las determinaciones reflejadas expresiones latinas, o porque el idioma materno no tiene ninguna expresión para ellas, o bien porque aun cuando las tenga, como en este caso, su expresión recuerda más lo inmediato, y la lengua extranjera, en cambio, más lo reflejado (…) El sentido y la expresión más exactos que el ser y la nada reciben puesto que desde ahora son momentos tienen que ser presentados (más adelante) en la consideración del ser determinado, como la unidad en la cual ellos son conservados. El ser es el ser y la nada es la nada sólo en su diversidad mutua; pero en su verdad, en su unidad han desaparecido como tales determinaciones y ahora son algo distinto. El ser y la nada son lo mismo y

Nótese que la definición de *Aufheben* dada por Hegel es, como puede observarse de la lectura de la nota al pie 10, una generalización lógica de tipo dialéctico de la cualidad que poseen el conjunto de fenómenos físicos conocidos como "momentos de fuerza", los cuales se inspiran la Estadística Matemática y la Estadística Aplicada para hablar de los Momentos de Probabilidad.

IV.VI. I.II. Las Probabilidades Condicionales y el Teorema de Bayes

El concepto de probabilidad condicional obedece una lógica similar, aunque menos general, que el concepto del Aufheben hegeliano. Sin embargo, el concepto de probabilidad condicional carecería de la suficiente profundidad filosófica para que su utilidad práctica fuese evidente hasta que el reverendo Thomas Bayes dijo "hágase la claridad analítica" (quizás la pidió, le fue dada y/o la encontró) y esta llegó, aunque como en promedio sucede con los genios revolucionarios, necesitó del nacimiento ni más ni menos que de un equivalente a escala aún más general (Pierre-Simon Laplace) para empezar a ser apreciado en una medida que hiciese justicia a su estatura intelectual, independientemente de sus alcances en contextos más generales y de su interpretación en tales contextos. El Teorema de Bayes es, por tanto, una forma de calcular e interpretar las probabilidades condicionales.

Así, la interpretación de las probabilidades condicionales se hará directamente en su forma más fundamental, desde el teorema de Bayes localizado en (DeGroot & Schervish, 2012, pág. 77):

$$\Pr(B_i|A) = \frac{\Pr(B_i)\Pr(A|B_i)}{\sum_{j=1}^{k} \Pr(B_j)\Pr\left(A\middle|B_j\right)} \qquad (17)$$

Es necesario comenzar por permitir que sea el mismo Bayes quien nos diga cómo se interpreta su teorema:

por este ser lo mismo, ya no son el ser y la nada y tienen una determinación diferente. Esta unidad constituye ahora su base; de donde ya no han de salir hacia el significado abstracto de ser y nada." (Hegel, 1968, págs. 97-98).

“DEFINITION (…) 5. The *probability of any event* is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon it’s happening.” (Bayes, 1763, pág. 376).

La interpretación que del Teorema de Bayes se hará en esta investigación estará orientada a la experimentación científica, por lo que se planteará desde el contexto de prueba de hipótesis[14]. Sin embargo, es fundamental conocer la interpretación objetiva, que es la interpretación óptima por cuanto lo objetivo contiene en última instancia a lo subjetivo (i.e., lo subjetivo es criatura siempre de lo objetivo[15]) del teorema en cuestión. Esta interpretación se establece en los siguientes términos: “Many applications of probability invoke a notion of probability that is objective in a logical sense: there is a fact of the matter as to what the probabilitie sare; if two agents disagree about a probability, at least one of them must be wrong. (Logical objectivity contrasts with the ontological sense of objectivity: probabilities are ontologically objective if they exist as entities or are reducible to existing entities, and are ontologically independent of mental or epistemological considerations.) For example, the probability that a patient’s breast cancer will recur after treatment apparently depends on features of the cancer, of the treatment, and of the patient. It is not simply a matter of personal opinion: if two prognostic probabilities differ, at least one of them must be wrong. A philosophical interpretation of probability should, if possible, yield a notion of probability that is suitably objective in this logical sense—otherwise, it is revising rather than faithfully interpreting probabilistic statements as they occur in these applications.” (Williamson, 2010, pág. 11).

[14] En los anexos de esta investigación se tratarán las probabilidades condicionales y la probabilidad total.

[15] Así, la definición de probabilidades aquí esbozada es compatible con los métodos utilizados por los subjetivistas (en cualquiera de sus niveles de radicalización), puesto que tales herramientas obedecen a un conjunto de axiomas (elaborados por Kolmogórov, como antes explicó), más no con la visión filosófica que orquesta sus espíritus científicos.

Por otro lado, la interpretación del Teorema de Bayes en el contexto de las pruebas de hipótesis puede hacerse siguiendo la lógica de (Russell, 2014). Ahí, el teorema mencionado toma la forma:

$$p(h|d) = \frac{p(d|h)\,p(h)}{p(d)} \qquad (18)$$

Como se menciona en la fuente citada, $p(h|d)$ puede interpretarse como qué tan verosímil es que nuestra hipótesis sea verdadera dada la evidencia científica disponible. Con ello, ahora resulta más intuitivo dar una explicación sobre el Teorema de Bayes, re-expresando las identidades (17) y (18) en una expresión diferente y de diferente ordenamiento de sus componentes:

$$\Pr(B_i|A) = \frac{\Pr(A|B_i)}{\Pr\,(A)}\Pr(B_i) \qquad (19)$$

En la expresión anterior, $\Pr(B_i|A)$ denota la probabilidad posterior, $\Pr(B_i)$ es la probabilidad a priori (conocida en teoría Bayesiana usualmente como *prior*, la cual representa el estado de conocimientos del investigador previo al encuentro de nueva evidencia relevante -la relevancia establecida por criterios que pueden variar según cada caso-) y $\frac{\Pr(A|B_i)}{\Pr\,(A)}$ es la razón (que en el numerador contiene al cociente de verosimilitud y en el denominador la probabilidad marginal asociada a una nueva y relevante evidencia encontrada -la probabilidad de cada "pieza de evidencia"-) e índice que mide la verosimilitud o confiabilidad asociada a esa nueva evidencia encontrada. Esto se entiende desde la lógica dialéctica-materialista del Teorema de Bayes y de las probabilidades condicionales en su verdad natural.

IV.VI. I. La Distribución Binomial Negativa II como Fenómeno En Sí

IV.VI. I. I. Conceptos Preliminares

La información presentada en esta sección referente puramente a los modelos jerárquicos y las mixturas de probabilidad se corresponde, salvo que se indique explícitamente lo contrario (mediante una cita bibliográfica) con el contenido expuesto por (Casella & Berger, 2002, págs. 163-166) interpolándole comentarios del autor de la presente investigación, los cuales serán indicados con una circunferencia diminuta cuya área tendrá su mismo color y este será negro. Expuesto lo anterior, debe comenzarse por decir:

- La diferencia entre un modelo jerárquico y una mixtura consiste en que el primero es una función matemática jerarquizada, i.e., es un instrumento de medición cuyo diseño responde al interés de la humanidad por medir fenómenos naturales y sociales para los que es significativamente relevante el hecho de que las etapas en que ocurre su desarrollo poseen una jerarquía bien-definida (según la teoría científica de referencia), mientras que las mixturas de probabilidad son funciones de distribución de probabilidad en que uno o más parámetros de la distribución de probabilidad (que puede ser univariante o multivariante, pero siempre implica probabilidad condicional y una modelación teórica del proceso de estimación de las distribuciones de probabilidad que responde a un diseño por etapas jerarquizadas) se modelan con base al comportamiento de alguna otra variable aleatoria, que condiciona implícitamente el comportamiento de la variable aleatoria estudiada. Estos modelos buscan capturar la noción de variable latente, variable implícita, variable oculta.
- Para comprender lo anterior debe introducirse sintéticamente el contexto de las variables latentes, el cual es el contexto de los modelos de ecuaciones estructurales (*SEM*, por sus siglas en inglés) y estas comúnmente en el contexto del meta-análisis en Bioestadística, Psicometría y Sociología. Según (Cheung, 2015, pág. 1), el meta-análisis tiene sus orígenes en Karl Pearson (1904) y su definición fue acuñada por Gene Glass en el contexto de la Psicología Educacional para representar "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings (…) (Glass 1976, p.3)"; por supuesto, el meta-análisis en su sentido más amplio no sólo es cuantitativo sino también

cualitativo y prueba de ello se encuentra, por ejemplo, en que el Axioma de Elección de Zermelo es un supuesto *ad hoc* cuya demostración no existe hasta la fecha[16] . Por otro lado, como se señala en (Kline, 2016, pág. 9), los SEM no hacen referencia a una única técnica estadística sino a una familia de procedimientos estadísticos relacionados entre sí; términos como *análisis estructural de covarianza*, *modelación estructural de covarianza* o *análisis de estructura de covarianza* son también utilizados en la literatura para clasificar esta familia de técnicas descrita, que a su vez describe una multiplicidad de técnicas "bajo una única etiqueta", gracias a rasgos comunes necesarios y suficientes que comparten en su estructura matemática para poder ser clasificados de tal forma, los cuales se sintetizan en que la estructura matemática general de los modelos estadísticos que pertenecen a la familia denominada SEM está diseñada para permitir una estimación especialmente robusta sobre las covarianzas de las variables aleatorias estudiadas por el investigador social o natural. Finalmente, "Latent variables in SEM generally correspond to **hypothetical constructs**, or explanatory entities resumed to rflecta continuum that is not directly observable. An example is the construct of *intelligence*. There is no single, definitive measure of intelligence. Instead, researchers use different types of observed variables such as task of verbal reasoning or memory capacity, to assess facets of intelligence." (Kline, 2016, pág. 12). La estructura teórica general de un modelo SEM puede verse en la figura presentada a continuación.

[16] Como se señala en (Billingsley, Probability and Measure, 2012, pág. 22), para espacios de probabilidad formados bajo consideraciones más complejas (que tienen que ver con cardinales transfinitos, i.e., conjuntos más grandes que el de los números naturales y cuya cardinalidad o tamaño se suele denotar por el símbolo $\aleph_0$, el cual se lee "Aleph Cero") es necesario asumir el Axioma de Elección. Por otro lado, en (Billingsley, Probability and Measure, 2012, pág. 47) se dice explícitamente que cuando los conjuntos son no numerables (es decir, no se puede establecer entre ellos y los naturales una biyección -una relación 1 a 1-) no podrían ser posibles las estimaciones estadísticas sin considerar el axioma en cuestión.

Figura 3

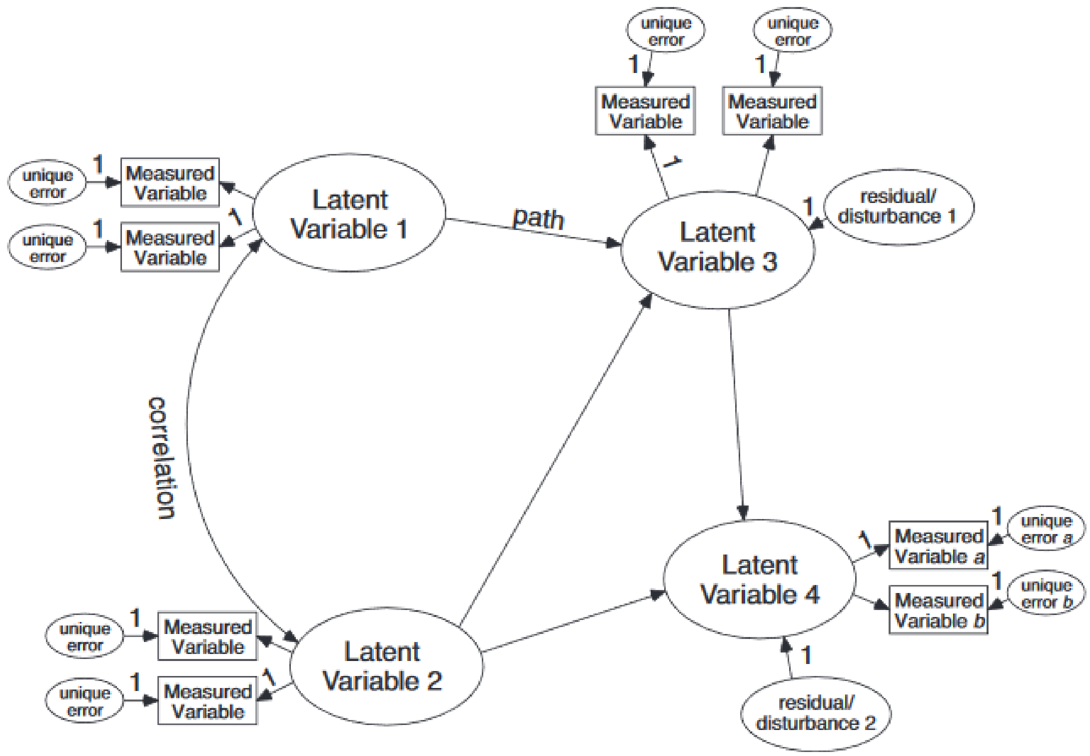


Figure 17.1 Full latent variable SEM model

Fuente: (Keith, 2019, pág. 390).

- Establecido lo anterior, es posible estudiar con la profundidad necesaria el tópico expuesto (Casella & Berger, 2002, págs. 163-166), que consiste en la relación matemática existente entre los modelos jerárquicos, las mixturas y cómo de la marginalización de una función jerárquica que expresa una mixtura de probabilidad compuesta por la distribución binomial, la distribución de Poisson y la distribución exponencial aparece la distribución binomial negativa 2.

IV.VI. I. II. Caso I de Modelo Jerárquico y Mixtura de Probabilidad: Binomial & Poisson

Quizás el caso más simple de modelo jerárquico de probabilidades es el presentado en esta sección, relacionado con la estimación de la probabilidad de que los huevos puestos por el ejemplar hembra de alguna especie de insecto que biológicamente está diseñada para poner un número sumamente grande de huevos, cada uno con probabilidad p de sobrevivir. Si se considera que cada uno de los huevos puestos por la hembra son independientes del otro (lo que implica que la probabilidad de sobrevivencia de uno de los huevos de sobrevivir no está relacionada con la del otro), puede verse cada huevo como un ensayo de Bernoulli.

Por tanto, si se toma como variable aleatoria las observaciones disponibles sobre el número de huevos que sobreviven (asumiendo que las observaciones con que se cuenta se comportan aleatoriamente -que es como asumir que sabemos de ellas menos de lo que creemos saber-) y se denota por X, puede estudiarse el número de sobrevivientes (que es aquí el fenómeno natural estudiado) mediante un modelo jerárquico en el que el número de sobrevivientes X sea una variable aleatoria cuyas probabilidades están modeladas por un funcional (una función de funciones) de distribuciones de probabilidad, en la que las probabilidades binomiales de sobrevivencia de los huevos p (si sobrevivió o no) a estimar tengan la característica de estar sujetas a considerar las probabilidades de que tras un determinado intervalo de tiempo haya sobrevivido un promedio λ de la cantidad $k-ésima$ de huevos antes descrita y no un número fijo que puede ser variable (un parámetro), con la finalidad, por ejemplo, de incrementar la incertidumbre del análisis y poder aproximarse a casos en que exista menos información disponible a la usual y/o de menor calidad a la usual. Un experimento de la naturaleza antes descrita, hace uso de un funcional que involucra múltiples (más de una, para este caso 2) funciones de probabilidad de la forma particular en que uno o más de sus parámetros se comporta a su vez como variable aleatoria (para este caso de estudio, una distribución binomial en que el parámetro n es transformado en una variable aleatoria que pertenece a la familia de distribuciones de Poisson de la forma definida en la identidad (15)). A esta clase especial de funcional se le conoce como *mixtura de probabilidades*. Así, este experimento toma la forma probabilística:

$$\begin{aligned} (X|Y) &\sim binomial(Y,p) \\ Y &\sim Poisson(\lambda) \end{aligned} \qquad (20)$$

Lo anterior expresa la distribución condicional de X dado $Y = y$, la cual es binomial.

La ventaja de los modelos jerárquicos es que fenómenos naturales y sociales complicados (vistos como procesos) pueden ser modelados por una secuencia de modelos relativamente simples ordenados en jerarquía (en que el proceso ocurre por etapas y que este hecho es relevante). Además, manipular modelos jerárquicos no es más difícil que hacer tales manipulaciones con distribuciones multivariantes condicionales y distribuciones marginales. De hecho, ya se explicó la conexión de las distribuciones multivariantes condicionales con los modelos jerárquicos. A continuación, se clarificará la relación entre los modelos jerárquicos y las mixturas de probabilidad.

- Si el lector revisa la fuente original, es decir, (Casella & Berger, 2002), podrá observar que la mayor parte del análisis realizado, salvo las estructuras matemáticas empleadas y el contexto académico sumamente general en que se explican estas estructuras (el ejemplo Bioestadístico empleado, más no en su especificidad), pertenecen exclusivamente a esta investigación. En este sentido, el ejemplo desarrollado en la forma en que se ha hecho aquí no sólo sirve para clarificar lo relativo a la fuente referida, sino también para poner en evidencia de forma más clara que sin el Teorema de Bayes las nociones de probabilidad condicional no terminarían de resultar intuitivas ni terminaría de tener una aplicabilidad teórica más general, así como tampoco las probabilidades en general. El anterior no es un modelo jerárquico bayesiano, por cuanto no utiliza la función de verosimilitud para calcular las probabilidades condicionales, sin embargo, a pesar de ello la sola existencia del Teorema de Bayes como concepto permite comprender de mejor forma las probabilidades condicionales en general, independientemente de la forma en que estas se calculen.

Una vez se ha definido el experimento y los fundamentos teóricos del mismo, solamente queda definir la estructura matemática de la variable aleatoria X:

$$P(X = x) = \sum_{y=0}^{\infty} P(X = x, Y = y)$$

$$\sum_{y=0}^{\infty} P(X = x|Y = y)P(\, Y = y)\,, lo\ cual\ es\ la\ definición\ de\ probabilidad\ condicional$$

$$P(X = x) = \sum_{y=x}^{\infty} \left[\binom{y}{x} p^{x}(1-p)^{y-x}\right]\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right] \qquad (21)$$

La identidad (21) es el tipo de modelo jerárquico que se conoce como mixtura de probabilidad, es decir, (21) es un funcional en que la función de distribución inmediata (la función de distribución binomial, para este caso) es de carácter condicional a su parámetro $n = y$, lo cual expresa que tal parámetro se comporta como una variable aleatoria cuya función de distribución de masa de probabilidad pertenece a la familia de distribuciones de Poisson[17]. Lo anterior es válido puesto que $(X|Y) = y$ es una distribución binomial (y, p) y $Y \sim Poisson(\lambda)$.

Ahora, si la expresión (21) se simplifica cancelando los términos semejantes y multiplicando por un 1 conveniente bajo la forma $\frac{\lambda^{x}}{\lambda^{x}}$, se obtiene:

$$P(X = x) = \frac{(\lambda p)^{x} e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{\big((1-p)\lambda\big)^{y-x}}{(y-x)!}$$

$$P(X = x) = \frac{(\lambda p)^{x} e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{\big((1-p)\lambda\big)^{t}}{t!}\,, en\ donde\ t = y - x$$

---

[17] Para determinar la forma del coeficiente binomial sólo se necesita sustituir n por y en la forma en que se planteó la función de masa binomial en la sección IV.III. No se hizo aquí por motivos de estética (para evitar la sobresaturación visual de la identidad).

$$P(X = x) = \frac{(\lambda p)^x e^{-\lambda}}{x!} e^{(1-p)\lambda} \quad (22)$$

$$P(X = x) = \frac{(\lambda p)^x}{x!} e^{-\lambda p} \quad (23)$$

En donde (22) es el *núcleo* de la distribución de Poisson[18]. Por tanto, es posible denotar este núcleo mediante $X \sim Poisson(\lambda p)$. Así, cualquier inferencia marginal (marginalización) sobre X es respecto a una distribución $Poisson(\lambda p)$, en la cual Y no desempeña ningún papel[19]. Lo anterior puede entenderse intuitivamente como llegar a la esencia de un fenómeno, pero recuérdese que la esencia desde una perspectiva dialéctica no se reduce a una característica, sino algo mucho más profundo[20], por lo que resulta posible generalizar esta definición teórica para una mixtura compuesta de $n - ésimos$ parámetros.

---

[18] El núcleo suele encontrarse en la literatura en inglés como *kernel*. Desde el punto de vista de la lengua inglesa, la palabra kernel es una modernización de la palabra *cyrnel* perteneciente al inglés antiguo, la cual significa "semilla" y, además, a nivel de la lexicología en habla inglesa empleada en Botánica, la palabra "core" se comprende como un contenedor de semillas); sin embargo, esta palabra tiene su origen en el idioma alemán y ahí su traducción es directamente *núcleo*. Según la Wikipedia en inglés (Wikipedia, 2020) tiene distintos significados según los campos de aplicación y según la Wikipedia en ruso (Википедия, 2020) es un concepto aplicado específicamente en Estadística Matemática, en Econometría, en Estadística Bayesiana, Estadística No-Paramétrica y la Teoría del Reconocimiento de Patrones verificando también (al igual que en versión en inglés) que su significado varía según su campo de aplicación. La Wikipedia en francés (Wikipedia, 2020) contribuye a terminar de clarificar la imagen cuando explica que los núcleos son parte de los denominados estimadores kernel (que sirven para estimar la función de densidad de probabilidad de una variable aleatoria, en la regresión paramétrica para estimar las esperanzas condicionales y en series de tiempo para estimar la densidad espectral. Por lo tanto, resulta evidente que nuestro caso de trabajo es a nivel de regresiones paramétricas en las que se estiman esperanzas condicionales (en otra investigación se extenderán estos resultados a la regresión binomial negativa 2) y en Estadística Bayesiana. En este contexto se habla del núcleo de una masa o de una densidad y es la forma que toma la masa o la densidad en la cual todos los factores que no son funciones de ninguna variable del dominio son omitidos de la masa o de la densidad, según corresponda. Este concepto busca capturar la intuición de "la esencia de la función", su *núcleo*.

[19] Distinta es la lectura que debe hacerse si se hablara de una *marginalización de* X (en este contexto X no desempeñaría ningún papel y la variable de relevancia sería Y. El lector puede notar que aquí se habla de una *marginalización sobre* X.

[20] La esencia puede ser comprendida como ese lugar intermedio entre dos opuestos, pero no intermedio en el sentido de que necesariamente ese lugar se encuentre en el centro de la trayectoria evolutiva del objeto; es ese lugar intermedio entre el punto de llegada y el punto de partida, la esencia está en el proceso de desarrollo de los objetos. En Hegel esos dos opuestos son el ser y el concepto, los cuales son en su sistema punto de llegada y punto de partida, respectivamente, mientras que en el sistema de Marx y Engels el ser es el punto de partida y el concepto es el punto de llegada, lo que

Introducir a Y en el modelo jerárquico fue principalmente para contribuir al entendimiento expositivo. Había una recompensa adicional en el sentido de que el parámetro de la distribución de X es el resultado de la combinación lineal de dos parámetros, cada uno relativamente simple de comprender.

Así, la respuesta a la pregunta inicialmente planteada es ahora fácil de estimar:

---

representa una inversión a la lógica de Hegel. Sin embargo, esa afirmación sería hueca si no se define qué domina en el *Aufheben*. Precisamente la dominancia es un aspecto analítico que caracteriza a Hegel y Marx como tan similares y a la vez tan diferentes. En ambos domina siempre el origen, que es fundamento del proceso y presenta dominancia en el resultado, sin embargo, es en la definición del origen en donde se diferencian Hegel y Marx. El segundo hace énfasis en esto en (Marx, 2007, págs. 14-18) al abordar la relación dinámica entre producción y consumo:

"Nada más simple, entonces, para un hegeliano que identificar producción y consumo. Y esto ocurrió no sólo en el caso de los ensayistas socialistas, sino también en el de economistas prosaicos como Say, p. ej., que piensan que si se considera a un pueblo su producción sería su consumo. O también a la humanidad in abstracto. Storch demostró el error de Say haciendo notar que un pueblo, p. ej., no consume simplemente su producción, sino que también crea medios de producción, etc., capital fijo, etc. (...) Además, considerar a la sociedad como un sujeto único es considerarla de un modo falso, especulativo- En un sujeto, producción y consumo aparecen como momentos de un acto. Lo que aquí más importa es hacer resaltar que si se consideran la producción y el consumo como actividades de un sujeto o de muchos individuos, ambas aparecen en cada caso como momentos de un proceso en el que la producción es el verdadero punto de partida y por ello también el momento predominante. El consumo como necesidad es el mismo momento interno de la actividad productiva. Pero esta última es el punto de partida de la realización y, por lo tanto, su factor predominante, el acto en el que todo el proceso vuelve a repetirse. El individuo produce un objeto y, consumiéndolo, retorna a sí mismo, pero como individuo productivo y que se reproduce a sí mismo. De este modo, el consumo aparece como momento de la producción. En la sociedad, en cambio, la relación entre el productor y el producto, una vez terminado este último, es exterior y el retorno del objeto al sujeto depende de las relaciones de éste con los otros individuos. No se apodera de él inmediatamente. Además, la aprobación inmediata del producto no es la finalidad del sujeto cuando produce en la sociedad. Entre el productor y los productos se interpone la distribución, que determina, mediante leyes sociales, la parte que le corresponde del mundo de los productos, interponiéndose por lo tanto entre la producción y el consumo (…) Ahora bien, ¿la distribución existe como una esfera autónoma junto a la producción y fuera de ella? (…) Aunque ésta aparezca como un supuesto para el nuevo período de producción, ella misma es a su vez producto de la producción, no solamente de la producción histórica en general, sino de la producción histórica determinada."

Así, la esencia es "un lugar" porque la esencia son relaciones de al menos un elemento con su entorno y/o consigo mismo (el caso "y" aplica para n-1 sistemas, el caso "o" sólo si se analiza el sistema universal, en que sólo existía la relación interna –puesto que no existe nada fuera del universo–), en que alguno tiene dominancia del otro en esa relación, con independencia de si es posible determinar esa dominancia o no. Entonces, son a esas relaciones a las que llamamos esenciales (o nucleares, que es lo mismo), porque su peso relativo es mayor en términos de las demás relaciones (tanto internas como externas) y ello significa además que los elementos involucrados (en estas relaciones esenciales) tendrán dominancia en general en su vinculación con elementos involucrados en relaciones no esenciales.

$$EX = \lambda p$$

Por tanto, en promedio, λp de los huevos sobrevivirá. Si se tuviese interés sólo en su media y se desease ignorar la distribución de las mismas, se podría recurrir a la utilización de las propiedades de las esperanzas condicionales para otro tipo de tratamiento teórico (que requeriría otro instrumental matemático). Además, en algunas ocasiones las estimaciones empíricas pueden ser simplificadas usando el siguiente teorema:

Sean X y Y dos variables aleatorias, entonces:

$$EX = E\big(E(X|Y)\big),$$

siempre que exista la esperanza matemática planteada.

Demostración

Sea $f(x,y)$ la función de distribución conjunta de distribución de la densidad de probabilidad de X y Y. Por definición se sabe que:

$$EX = E\big(E(X|Y)\big) = \iint xf(x,y)dxdy = \int \left[\int xf(x|y)dx\right] f_Y(y)dy,$$

En donde $f(x|y)$ y $f_Y(y)$ son las densidades condicionales de X dado $Y = y$ y la distribución marginal de Y, respectivamente. Nótese que la integral interior es la esperanza condicional[21] $E(X|y)$, que es la esperanza del parámetro de la distribución de Poisson (λ) por lo que se tiene que:

$$EX = E\big(E(X|Y)\big) = \int E(X|y)f_Y(y)dx = \ p\lambda$$

[21] Recuérdese que las integrales son momentos de probabilidad o esperanzas matemáticas (o medias aritméticas), en este caso momentos condicionales, esperanzas matemáticas condicionales o medias aritméticas condicionales.

Para la demostración el caso discreto sólo se debe sustituir las integrales por sumas y no agrega claridad analítica, por lo que se omitirán[22].

Establecido lo anterior, pueden consolidarse tales resultados de la forma siguiente:

$$EX = E\big(E(X|Y)\big)$$

$$= E(pY), \qquad puesto\ que\ (X|Y) \sim binomial\ (Y, p)$$

$$= p\lambda, \qquad puesto\ que\ Y \sim Poisson(\lambda)$$

El término *distribución mixta* o *mixtura de probabilidad* en el título de esta sección hace referencia a una distribución surgida de una estructura jerárquica. Aunque no existe una definición estandarizada para este término, la definición anterior es la que mayor aceptación posee actualmente en la comunidad científica internacional.

Además, la riqueza filosófica de este enfoque no se agota ahí, por el contrario. En ese sentido, $E\big(E(X|Y)\big)$ es la media de una media, el promedio de un promedio. Quizás sin especificaciones topológicas de algún tipo no parece ser posible formarse una idea intuitiva y de capacidad aplicativa materializable sobre qué es la media de una media (que es otra forma de ver la probabilidad condicional, equivalente a todas las anteriores), pero si el espacio muestral se generaliza a n-ésimas dimensiones es posible observar en los resultados anteriores un patrón algorítmico iterativo que es la base de procesos numéricos realizados mediante el uso intensivo de algoritmos iterativos computacionales, como ocurre con la familia de modelos lineales generalizados (que utilizan el algoritmo conocido como *Iteratively Reweighted Least Squares* -IRLS-) o el Método Monte Carlo descubierto por

---

[22] Como señalan (Casella & Berger, 2002, pág. 164), $EX = E\big(E(X|Y)\big)$ puede categorizarse como un "abuso de notación" puesto que se ha utilizado la E para simbolizar diferentes esperanzas matemáticas en la misma ecuación. Por ello, es necesario enfatizar que la E del lado izquierdo de la identidad es la esperanza o momento con respecto a la distribución marginal de X; mientras que la E exterior del lado derecho es la esperanza o momento con respecto a la distribución marginal de Y y la E interior del lado derecho es la esperanza o momento con respecto a la distribución condicional de $(X|Y)$. Sin embargo, como señalan los autores "However, there is really no cause for confusion because these interpretations are the only ones that the symbol "E" can take!".

Ulam y Neumann en Los Álamos. Lo anterior no representa otra sino distintos tipos de procesos de convergencia hacia la media.

IV.VI. I. III. Caso II de Modelo Jerárquico y Mixtura de Probabilidad: Binomial, Poisson & Exponencial

La presente sección se constituye como una generalización inmediata de la sección anterior, en la que en lugar de una madre insecto existen un número sumamente grande de madres y cada una de las madres es escogida al azar. Si en el escenario antes descrito un investigador estuviese interesado en conocer la media aritmética de los sobrevivientes (i.e., el número promedio de sobrevivientes de repetirse multiplicidad de veces el experimento), pero en la investigación no está aún lo suficientemente claro con base en la evidencia obtenida y disponible en general (ni de los experimentos Bioestadísticos ni del marco teórico de la Biología) de que el número de huevos puestos siga la misma distribución de Poisson para cada madre, el siguiente modelo jerárquico de tipo mixtura de probabilidad puede ser más adecuado.

Sea X el número de sobrevivientes en una camada, entonces:

$$\begin{gathered}(X|Y) \sim binomial\ (Y, p) \\ (Y|\Lambda) \sim Poisson(\Lambda) \\ \Lambda \sim exponencial(\beta)\end{gathered} \qquad (24)$$

En la jerarquía anterior la última etapa (que corresponde a la distribución exponencial) modela a variabilidad entre diferentes madres.

La media de X puede calcularse fácilmente como:

$$EX = E\big(E(X|Y)\big)$$

$$E(pY), \qquad como\ se\ hizo\ en\ la\ sección\ anterior$$

$$= E\big(E(pY|\Lambda)\big)$$

$$= E(p\Lambda)$$

$$= p\beta, \qquad esperanza\ exponencial$$

Finalizando así el proceso de cálculo de la esperanza de la mixtura.

En el ejemplo anterior se utilizó un modelo sutilmente diferente al utilizado en la sección anterior en el sentido de que dos variables aleatorias eran discretas y una continua. Usar este tipo de modelos no debería de presentar problemas. Se puede definir la densidad conjunta $f(x, y, \lambda)$, las densidades condicionales $f(x|y)$ y $f(x|y, \lambda)$, etc., así como las densidades marginales $f(x)$, $f(x, y)$, etc., como se hizo antes. Simplemente se debe comprender que cuando las probabilidades o los momentos de probabilidad son calculados, las variables discretas son sumadas y las variables continuas son integradas.

Así, puede evidenciarse que el modelo de tres etapas descrito anteriormente puede re-describirse como un modelo de dos etapas combinando las últimas dos etapas[23]. Sea $(Y|\Lambda) \sim Poisson(\Lambda)$ y $\Lambda \sim exponential(\beta)$, entonces:

$$P(Y = y) = P(Y = y, 0 < \Lambda < \infty)$$

$$P(Y = y) = \int_0^\infty f(y, \lambda) d\lambda$$

$$P(Y = y) = \int_0^\infty f(y|\lambda) f(\lambda) d\lambda$$

$$P(Y = y) = \int_0^\infty \left[\frac{e^{-\lambda}\lambda^y}{y!}\right] \frac{1}{\beta} e^{-\lambda-\beta} d\lambda$$

---

[23] Por supuesto, esto sólo sería válido si lo es en términos del marco teórico, lógico y del acervo experimental de la ciencia en la que se esté trabajando, lo cual puede requerir para su determinación un análisis riguroso de carácter teórico y aplicado que certifique que analizar el fenómeno desde otra estructura matemática es no solamente su equivalente matemático sino también su equivalente desde tales marco teórico-científico, lógica de la ciencia y acervo experimental.

$$P(Y=y)=\frac{1}{\beta y!}\int_0^\infty \lambda^y e^{-\lambda\left(1+\beta^{-1}\right)}d\lambda \qquad (25)$$

Merece la pena hacer una pausa para comentar que (25) es la densidad de probabilidad gamma.

$$P(Y=y)=\frac{1}{\beta y!}\Gamma(y+1)\left(\frac{1}{1+\beta^{-1}}\right)^{y+1}$$

$$P(Y=y)=\frac{1}{(1+B)}\left(\frac{1}{1+\beta^{-1}}\right)^{y} \qquad (26)$$

La identidad (26) se puede re-expresar como:

$$P(Y=y)=\binom{r+y-1}{y}p^r(1-p)^y, \qquad y=0,1,\ldots. \qquad (27)$$

Así, (27) es equivalente a la forma (3.2.10) que exponen (Casella & Berger, 2002, pág. 95), como se señala en (Casella & Berger, 2002, pág. 192). Así, la identidad (27) expresa la distribución de la masa de probabilidad de la *Distribución Binomial Negativa II* que constituye el punto de llegada en esta investigación[24].

Finalmente, puede entenderse así la analogía entre el estudio de la distribución binomial negativa 2 como el resultado de un proceso con el Ser En Sí, así como también su estudio particular o marginal como Ser Para Sí.

IV.VI. II. La Distribución Binomial Negativa II como Fenómeno Para Sí

En la sección anterior se desarrollaron sistemáticamente las mixturas de distribución de probabilidad para el caso de un modelo jerárquico de 2 etapas (Binomial-Poisson) y de 3 etapas (Binomial-Poisson-Exponencial), para luego comprimir el modelo de 3 etapas en uno de 2 etapas (Poisson-Exponencial) y tras

[24] Más adelante se presentará la misma función, aunque en lugar de y aparecerá x. Esto no debe confundirse con que en algunas literaturas se hace un despeje de la variable independiente, sino que debe entenderse simplemente como el uso de otra letra, tal y como se hace ampliamente en la literatura disponible.

marginalizar este último obtener la distribución binomial negativa 2 (NB2 de ahora en adelante, por sus siglas en inglés).

La ruta de trabajo de esta sección consiste en un estudio superficialmente general de la distribución marginalizada anteriormente, un estudio breve de sus dos versiones más conocidas y un estudio con relativa profundidad sobre su segunda versión, que es el punto de llegada de esta investigación.

Como señala (Hilbe, 2011, pág. xi)[25], la distribución binomial 2 tiene como finalidad ser una herramienta que pueda modelar conteo de datos y posee un espectro amplio de versiones de sí misma[26], sin embargo, es la NB2 la que ha pasado a ser conocida como *distribución binomial negativa de uso estándar*, lo cual se debe a que progresivamente fue convirtiéndose en uno de los métodos principales para analizar modelos de respuesta por conteo; sin embargo, como señala el autor referido en el mismo lugar, relativamente pocos investigadores y estadísticos aplicados están familiarizados con las variedades disponibles de modelos de la distribución binomial negativa, así como tampoco con la mejor forma de incorporar tales modelos a su ruta de investigación. El modelo de regresión de Poisson, tradicionalmente considerado el modelo de conteo básico[27], puede verse como una instancia de la distribución NB2, específicamente es una distribución NB2 con parámetro de heterogeneidad igual a cero[28]. Es por ello que señala Hilbe

[25] Joseph Hilbe es investigador del departamento de propulsión a chorro de la NASA, profesor del California Institute of Technology y profesor de la Arizona State University.
[26] Cada una sigue su propia ruta, como puede verificarse en la investigación (Swat, Grenon, & Wimalaratne, 2017).
[27] Hilbe hace referencia aquí a los modelos de regresión con variables latentes.
[28] El lector puede consultar (Hilbe, 2011, pág. 319) para más detalle en el contexto de regresión, que se escapa del foco de esta investigación. Sin embargo, "It is important to understand that the traditional negative binomial model can be estimated using a standard maximum likelihood function, or it can be estimated as a member of the family of generalized linear models (GLM). A

que quizás la NB2 es el modelo de conteo más general y quizás también el más representativo entre todos los modelos de conteo utilizados en la investigación diaria que se suscita en la actualidad.

Como relata (Hilbe, 2011, pág. 5), los orígenes históricos de la NB pueden conocerse mediante un reporte realizado por Isaac Todhunter titulado *History of Mathematical Theory of Probability from the Time of Pascal to that of Laplace* publicado en 1865. Ahí, se dice que Pierre de Montmort, aristócrata y matemático francés del siglo XVIII, discípulo de Nicolas Malebranche (filósofo y teólogo francés), la mencionó en 1713 en el contexto de la estimación de un número de fallas y antes del $k - ésimo$ éxito en una serie de ensayos binarios, lo que relata cómo la NB nace como NB2. Como se señala en al fuente referida sobre la familia NB: "Little was done with either the earliest definition of negative binomial as derived by Montmort, or with Poisson's distribution for describing count data, until the early twentieth century. Building on the work originating with Gauss (1823), who developed the normal, or Gaussian, distribution, upon which ordinary least squares (OLS) regression is based, the negative binomial was again derived by William Gosset, under his pen name, Student, in 1907 while working under Karl Pearson at his Biometric Laboratory in London (Student, 1907). In the first paper he wrote while at the laboratory, he derived the negative binomial while investigating the sampling error involved in the counting of yeast cells with a haemocytometer. The paper was published in Biometrika, and appeared a year earlier than his well-regarded papers on the sampling error of the mean and correlation coefficient (Jain, 1959). However, G. Udny Yule is generally, but arguably, credited with formulating the first negative binomial distribution based on a 1910 article dealing with the distribution of the number of deaths that would occur as a result of being exposed to a disease (i.e. how many deaths occur given a certain number of exposures). This formulation stems from what is called inverse binomial sampling.

negative binomialmodel is aGLMonly if its heterogeneity parameter is entered into the generalized linear models algorithm as a constant." (Hilbe, 2011, pág. 3).

Later Greenwood and Yule (1920) derived the negative binomial distribution as the probability of observing y failures before the rth success in a series of Bernoulli trials, replicating in a more sophisticated manner the work of Montmort. Three years later the contagion or mixture concept of the negative binomial originated with Eggenberger and Polya (1923). They conceived of the negative binomial as a compound Poisson distribution by holding the Poisson parameter, λ, as a random variable having a gamma distribution. This was the first derivation of the negative binomial as a Poisson–gamma mixture distribution. The article also is the first to demonstrate that the Poisson parameter varies proportionally to the chi2 distribution with 2 degrees of freedom." (Hilbe, 2011, págs. 5-6)

En una de las investigaciones citadas por (Hilbe, 2011), específicamente en (Greenwood & Yule, 1920), se presenta y estudia el escenario en el que se realiza un *ajuste de curvas*[29] para superponer los polígonos que constituyen las gráficas de la distribución binomial y la distribución binomial negativa 2 con el fin de mostrar su identidad y relación bajo determinadas condiciones especificadas por los investigadores citados[30], tal y como se presenta a continuación:

[29] Proceso conocido geométricamente como *ajuste de curvas*.

[30] El lector puede verificar que la NB que aparece en (Greenwood & Yule, 1920), puesto que en (Agarwal, Bajorski, Farnsworth, Marengo, & Qian, 2017, pág. 2) se puede leer que "There is much current interest in compounding or mixing distributions and their applications. Indeed, the early history of statistics was greatly concerned with the problem [3]. The work by Greenwood and Yule (…) in the more modern era has been followed up with new results and extensive applications (…) including many based on the Poisson distribution because of its centrality in statistical analysis and probability modeling (…) The present derivations supply new insights into the structure of this type of modeling by revealing how compounded Poisson variables produce a negative binomial distribution.", por lo que junto con lo expuesto por los autores referidos en la página 2 del mismo documento, puede verificarse que en (Greenwood & Yule, 1920) se habla de la NB2.

Figura 4

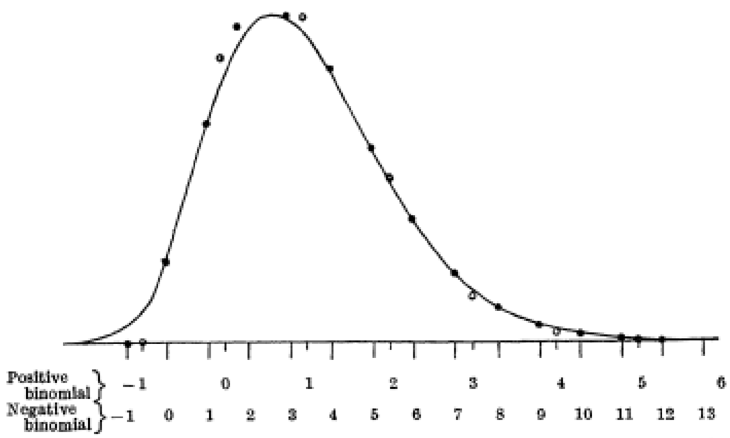


Diagram showing binomial curve fitted, by the slope relation, to both a positive binomial (hollow circles) and a negative binomial (circles blocked in).

Fuente: (Greenwood & Yule, 1920, pág. 279).

Además, la sección de la investigación de Stanley Gossens referida por Hilbe, la cual se encuentra en (Student, 1907, págs. 355-360), se obtiene también una de las versiones de la NB, como puede observarse a continuación:

Figura 5

The actual distributions of cells are given below, and compared with those calculated on the supposition that they are random samples from a population following the law which we have investigated: the probability P of a worse fit occurring by chance is then found.

I. Mean $= \cdot 6825 : \mu_2 = \cdot 8117 : \mu_3 = 1 \cdot 0876$.

Containing	0	1	2	3	4	5 cells
Actual	213	128	37	18	3	1
Calculated	202	138	47	11	1·84	·24
					2	

Whence $\chi^2 = 9 \cdot 92$ and $P = \cdot 04$.

Best fitting binomial $(1 \cdot 1893 - \cdot 1893)^{-3 \cdot 6054} \times 400$ for which $P = \cdot 52$.

II. Mean $= 1 \cdot 3225 : \mu_2 = 1 \cdot 2835\ \mu_3 : = 1 \cdot 3574$.

	0	1	2	3	4	5	6
Actual	103	143	98	42	8	4	2
Calculated	106	141	93	41	14	4	1

Whence $\chi^2 = 3 \cdot 98$ and $P = \cdot 68$.

Best fitting binomial $(\cdot 97051 + \cdot 02949)^{46 \cdot 2084} \times 400$ for which $P = \cdot 72$.

III. Mean $= 1 \cdot 80 : \mu_2 = 1 \cdot 96 : \mu_3 = 2 \cdot 529$.

	0	1	2	3	4	5	6	7	8	9
Actual	75	103	121	54	30	13	2	1	0	1
Calculated	66	119	107	64	29	10	3		1	

Whence $\chi^2 = 9 \cdot 03$ and $P = \cdot 25$.

Best fitting binomial $(1 \cdot 0889 - \cdot 0889)^{-20 \cdot 2473} \times 400$ for which $P = \cdot 37$.

IV. Mean $= 4 \cdot 68 : \mu_2 = 4 \cdot 46 : \mu_3 = 4 \cdot 98$.

	0	1	2	3	4	5	6	7	8	9	10	11	12
Actual	0	20	43	53	86	70	54	37	18	10	5	2	2
Calculated	4	17	41	63	74	70	54	36	21	11	5	2	1

Whence $\chi^2 = 9 \cdot 72$ and $P = \cdot 64$.

Best fitting binomial $(\cdot 9525 + \cdot 0475)^{98 \cdot 53} \times 400$ for which $P = \cdot 68$.

Fuente: (Student, 1907, págs. 356-357).

Puede observarse que la estimación III corresponde a una distribución cuya varianza es superior a su media[31] y que el ajuste del polígono binomial se realizó mediante una expansión negativa. Así, puede concluirse que Gosset obtuvo ahí la

[31] El lector debe recordar que anteriormente se definió esta distribución como una de las características distintivas de la NB que le permiten ser y ser vista como una generalización de la distribución de Poisson.

distribución binomial negativa, que a juzgar por las series de potencias desarrolladas y por lo visto en el lugar referido por el autor en cuestión, así como también por lo visto en la presente investigación, parece ser que se trata de la distribución binomial negativa 1[32].

En este sentido, el lector podrá encontrar en los anexos de esta investigación una explicación y un caso aplicado simple de la NB1, puesto que su utilidad no es menor; sin embargo, en el resto del cuerpo principal de esta investigación se hará énfasis en la NB2.

Como se mostró anteriormente en la presente investigación, el número de éxitos (con probabilidad p) en una sucesión de $n-ésimos$ los ensayos de Bernoulli siguen una distribución binomial con parámetros n y p. Sin embargo, ¿qué ocurriría si desea estudiar tales ensayos (que representan observaciones y eventos, de un experimento compuesto -la distribución de Bernoulli puede comprenderse como un experimento simple-) hasta que se obtuviese algún número deseado de éxitos?, o, dicho de forma más general, ¿cómo modelar el estudio científico de un fenómeno natural o social que se desea abordar teóricamente en términos del curso del desarrollo, que implica la manifestación de sus múltiples rasgos o características de tal o cual forma y de tal o cual importancia específica, hasta que este desarrollo manifestara alguno de esos tales o cuales rasgos o características definidos por el investigador según las necesidades concretas de la investigación? Los grandes personajes de la Estadística Matemática y la Estadística Aplicada con quienes tanto el lector como yo hemos tenido el placer de conversar a lo largo de

[32] Debido a que las formas de los desarrollos matemáticos han cambiado profundamente desde que Gosset escribió su investigación en los laboratorios bioquímicos de Karl Pearson entre 1907 y 2008 no es fácil establecer cuando dos objetos matemáticos son equivalentes debido a toda la evolución que ha tenido la Estadística Matemática desde que fue fundada por Karl Pearson hace alrededor de 100 años. Si el lector echa un vistazo rápido a las investigaciones de Gosset se dará cuenta que las matemáticas involucradas son similares a las contenidas en las investigaciones de Karl Pearson en cuanto a estilo, y esto es así porque Karl Pearson fue decisivo en el robustecimiento matemático de las investigaciones de Gossens.

esta investigación se encargaron de resolver de distintas formas la segunda pregunta realizada anteriormente.

Finalmente, se plantearán los momentos de la distribución binomial negativa 2, su función generadora de momentos y su función característica, sobre la cual se hablará con un poco más de detalle en los anexos.

Estos constructos teóricos (los objetos matemáticos bajo la etiqueta de *momentos de probabilidad*) son las $n - \acute{e}simas$ sub-localizaciones[33] dentro de un *Campo de Probabilidad*[34] que proporcionan a la distribución de probabilidad su forma, escala y otros aspectos de su localización sub-regional específica (en la $sub - esfera$ de la $n - esfera$ en que la distribución está definida) de forma completa dentro del campo de probabilidad al que pertenece la distribución de probabilidad estudiada. Los momentos descritos a continuación son tomados de (Bagui & Mehra, 2019, pág. 44). Los momentos expuestos a continuación son tomados de (Bagui & Mehra, 2019, pág. 44), mientras que la función generatriz de momentos y función característica pueden encontrarse en (Walck, 1996, pág. 102) y en (Wikipedia, 2020):

Primer Momento de Probabilidad (Nombre: Media Artimética)

$$E(X) = \frac{rq}{p}, \qquad donde\ q = 1 - p$$

---

[33] Según (Wikipedia, 2020), los grados de libertad pueden definirse como "El número de formas independientes por las que un sistema dinámico puede moverse (desarrollarse), sin violar ninguna restricción impuesta (en el modelo), se denomina número de grados de libertad. En otras palabras, el número de grados de libertad se puede definir como el número mínimo de coordenadas independientes que pueden especificar la posición del sistema por completo", por lo que las coordenadas del campo de probabilidad a nivel de su estructura (la base canónica, conjunto generador o conjunto de vectores linealmente independientes que generan el espacio lineal), i.e., los grados de libertad, están dadas; por ello es que se regionalizan aquí las coordenadas, puesto que como señala la fuente antes citada "Si bien los libros de texto introductorios pueden introducir grados de libertad como parámetros de distribución o mediante pruebas de hipótesis, es lao geometría subyacente la que define los grados de libertad y es fundamental para una comprensión adecuada del concepto."

[34] Entiéndase por ellos como los define (Kolmogórov, 1956, pág. 2).

Segundo Momento de Probabilidad (Nombre: Varianza)

$$E(X^2) = \frac{rq}{p^2}, \qquad donde\ q = 1 - p$$

Aplicando a $E(X^2)$ un poco de álgebra puede transformarse de la siguiente forma:

$$\sigma^2 = \frac{\mu}{p} = \frac{\mu(p+q)}{p} = \mu + \frac{\mu^2}{r}, \qquad donde\ q = 1 - p$$

Puede verificarse de la identidad anterior que la distribución binomial negativa 2 tiene la característica de que su varianza puede ser más grande que su media y esto es precisamente el hecho que le permite erigirse como una alternativa más flexible que la distribución de Poisson, es decir, una generalización en el contexto definido.

Tercer Momento de Probabilidad (Nombre: Coeficiente de Asimetría)

$$E(X^3) = \frac{2 - p}{\sqrt{r(1-p)}}$$

Cuarto Momento de Probabilidad (Nombre: Curtosis[35])

$$E(X^4) = \frac{6}{r} + \frac{p^2}{r(1-p)}$$

Función Generatriz de Momentos

$$MGF = \left(\frac{1-p}{1-pe^t}\right)^r, \qquad para\ todo\ t < -\log(p)$$

Función Generatriz de Probabilidades

$$PGF = \left(\frac{1-p}{1-pz}\right)^r, \qquad para\ todo\ |z| < \frac{1}{p}$$

Función Característica

$$CF = \left(\frac{1-p}{1-pe^{it}}\right)^r, \qquad en\ donde\ t \in \mathbb{R}$$

[35] Del griego κυρτός. Esa es la razón por la que en algunos lugares se encuentra como “kurtosis” y en el idioma inglés se escribe con la letra inicial k.

Lo anterior[36] puede resumirse en las siguientes imágenes, en las gráficas de distribución de masa binomial negativa son construidas mediante la asignación de diferentes valores para la variable y parámetros de interés:

Figura 6

Moments

Mean:	$\frac{r(1-p)}{p}$
Variance:	$\frac{r(1-p)}{p^2}$
Mode:	The largest integer $\leq \frac{(r-1)(1-p)}{p}$.
Mean Deviation:	$2u\binom{r+u-1}{u}(1-p)^u p^{r-1}$, where u is the smallest integer greater than the mean. [Kamat 1965]
Coefficient of Variation:	$\frac{1}{\sqrt{r(1-p)}}$
Coefficient of Skewness:	$\frac{2-p}{\sqrt{r(1-p)}}$
Coefficient of Kurtosis:	$3 + \frac{6}{r} + \frac{p^2}{r(1-p)}$
Central Moments:	$\mu_{k+1} = q\left(\frac{\partial \mu_k}{\partial q} + \frac{kr}{p^2}\mu_{k-1}\right)$, where $q = 1-p$, $\mu_0 = 1$ and $\mu_1 = 0$.

Moment Generating Function:	$p^r(1-qe^t)^{-r}$
Probability Generating Function:	$p^r(1-qt)^{-r}$

Fuente: (Krishnamoorthy, 2006, págs. 98-99).

[36] En los anexos se expondrá la estructura matemática de los primeros dos momentos muestrales de la NB2.

Figura 7

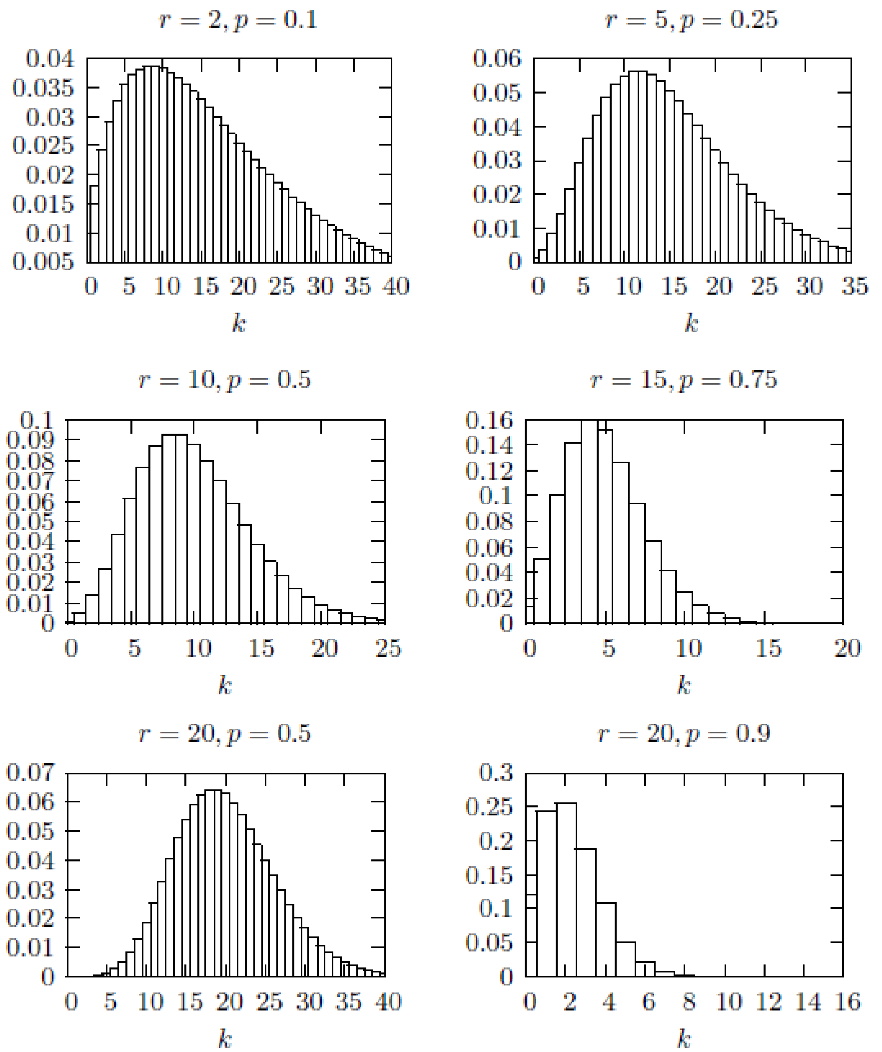


Figure 7.1 Negative Binomial Probability Mass Functions; k is the Number of Failures until the rth Success

Fuente: (Krishnamoorthy, 2006, pág. 99).

Figura 8

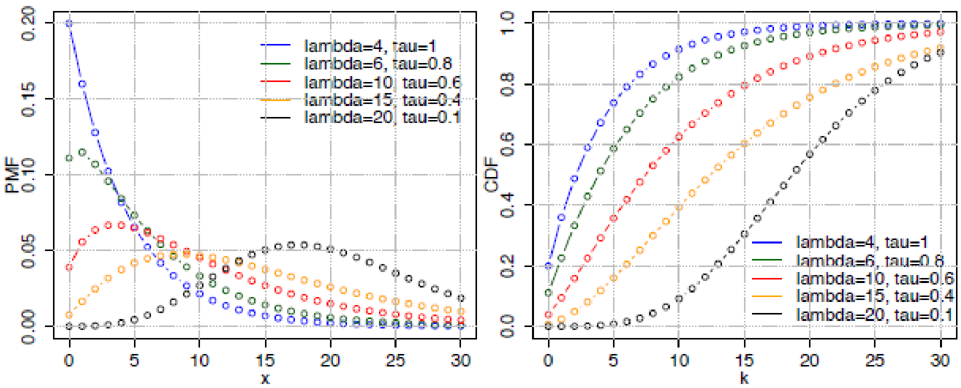


Figure A.83: NegativeBinomial2 distribution plotted using the provided R code.

Fuente: (Swat, Grenon, & Wimalaratne, 2017, pág. 197).

La capacidad de resolver problemas aplicados latentes en el instrumental anterior puede verse, por ejemplo, como se señala en (DeGroot & Schervish, 2012, pág. 297), para el caso en que se desee monitorear[37] algún equipo, se va a suponer aquí que es capital fijo perteneciente al Sector I (Sector Productor de Medios de Producción), con la finalidad de determinar los puntos del desarrollo de su vida útil en que necesita mantenimiento. La estimación del número de puntos[38] o eventos de interés a observar en dicha trayectoria evolutiva del sistema previo a que la alarma del equipo de monitoreo notifique al operario de tal monitor un número fijo de veces que el equipo requiere mantenimiento se puede modelar con la distribución binomial negativa 2.

Así, supóngase que el capital fijo definido anteriormente produce partes que pueden ser óptimas o defectuosas. Sea $X_i = 1$ si la $i - ésima$ parte es defectuosa (puesto que se quiere encontrar el punto en que la alarma notifique necesidad de

[37] Mediante la utilización de algún equipo de monitoreo especializado para el caso.
[38] Los puntos de la trayectoria de un sistema (de ecuaciones diferenciales o en diferencias) son los instantes temporales que conforman su trayectoria evolutiva (conocida como evolución del sistema, que en este caso es la evolución en el sentido de su vida útil hasta alcanzar un valor de descarte contable). Por ejemplo, si una computadora tiene una vida útil contable de 5 años y se desea monitorear su vida útil en unidades temporales anuales, los puntos serían 5, en donde cada punto representaría 1 año de los 5 años de vida útil contable en total.

mantenimiento) y sea $X_i = 0$ en cualquier otro lado. Asumiendo que tanto las partes óptimas como las defectuosas son independientes (es decir, la producción de una pieza defectuosa no afecta la producción de otra, por lo que sus probabilidades de ocurrencia no poseen relación) y cada una de ellas con probabilidad $P(X_i = 1) = p$ para todo $i - ésima$ parte defectuosa.

Esta secuencia de ensayos de Bernoulli implícita en la NB2 puede ser infinita, en donde cada una de las opciones son definidas en términos de éxito o fracaso (en este caso, los éxitos son cada punto en el que el capital fijo requiere mantenimiento a causa de la detección de la producción de una pieza defectuosa) con probabilidad de éxito p. En esta sección se estudiará la distribución de probabilidad del número de fracasos que ocurrirán antes que ocurran exactamente r éxitos en el experimento definidos en los términos expuestos anteriormente, en dónde r es cualquier entero no negativo.

Así, en un experimento diseñado bajo tales especificaciones, es posible utilizar la distribución de masa de probabilidad binomial negativa, en la que X representa el número de fracasos antes de que el $r - ésimo$ éxito ocurra, que adopta la siguiente forma funcional:

$$f(x|r,p) = \begin{bmatrix} \binom{r+x-1}{1} p^r (1-p)^x \\ 0, \quad en\ otro\ caso \end{bmatrix}, \qquad para\ todo\ x = 0, 1, 2, \ldots, \tag{28}$$

Nótese que (28) pertenece a (DeGroot & Schervish, 2012, pág. 297), mientras que (27) al final de la sección anterior pertenece a (Casella & Berger, 2002, pág. 95), sin embargo, a todas luces son equivalentes. Además, nótese que la expansión del coeficiente binomial en (28) no aparece en ninguna de las fuentes referidas. Sin embargo, puede extraerse esta información tanto de (Mittelhammer, 2013, pág. 182), específicamente de la línea #6 y #7 del primer párrafo, así como también de (Wikipedia, 2020). Ahí puede verificarse que dicho coeficiente permitiría re-expresar (28) en los siguientes términos:

$$f(x|r,p) = \begin{bmatrix} \left(\frac{(x+r-1)!}{(r-1)!\,(x)!}\right) p^r (1-p)^x \end{bmatrix}, \qquad para\ todo\ x = 0,1,2,\ldots, \tag{29}$$
$$0, \qquad en\ otro\ caso$$

La distribución binomial negativa debe comprenderse, en general, en términos del número de ensayos de Bernoulli (experimentos simples) requerido para obtener algún número fijo de éxitos, tal como se observa en (Casella & Berger, 2002, pág. 95), así como también en (Mittelhammer, 2013, págs. 181-183) y en (Wackerly, Mendelhall III, & Scheaffer, 2010, págs. 121-123), que es la distribución binomial negativa 1 que se tratará en los anexos; por ello debe señalarse, como observa (Cohen & Cohen, 2008, pág. 166) que "The general definition of the negative binomial does not require that n be a nonnegative integer. However, because we are interested in the physical interpretation of the density, we will focus our attention on positive integers". Sin embargo, tanto (28) como (29) pertenecen a la distribución binomial negativa 2, que se obtiene redefiniendo la necesidad de investigación que la herramienta NB1 resuelve en términos del **número de ensayos de Bernoulli, que se definen como fracasos**, requerido para obtener obtener un número fijo de éxitos, con lo que se obtiene la distribución binomial negativa 2. Sobre la distribución binomial negativa 1 se hablará en los anexos con mayor detalle.

Como se explicó anteriormente, la razón por la que aquí se ha hecho énfasis analítico en NB2 en lugar de NB1 se encuentra en la facilidad de alcanzar mayor riqueza analítica al estudiar la NB en términos de un cuerpo teórico mucho más general que ella, así como también por sus orígenes históricos, que es tan viejo como la Estadística, tal y como se verá a continuación; esta investigación puede tener alguna utilidad en allanar el camino para una réplica de sí en el que la distribución binomial negativa de estudio fuese la NB1, sin embargo, ello probablemente podría implicaría un esfuerzo mayor que el realizado en esta investigación concerniente a la recopilación de fuentes históricas al respecto, lo cual tiene que ver en alguna medida con que la NB2 facilita las estimaciones en relación a sus pares, aunque la NB1 es la NB que responde a una pregunta más general, al menos en comparación con la NB2.

Si el lector consulta (Ibe, 2013, pág. 19), notará que la distribución de Pascal que el autor presenta es equivalente a la forma de la NB (Wackerly, Mendelhall III, & Scheaffer, 2010, pág. 122) y (Mittelhammer, 2013, pág. 182) de la NB1, que aparece junto con la NB2 tanto en (Walck, 1996, pág. 102) como en (Casella & Berger, 2002, pág. 95)[39], específicamente en el lugar en el que inicia su exposición sobre la NB, que a su vez se corrobora según (Cook, 2009, pág. 1) que es la forma funcional de la NB etiquetada como NB1; sin embargo, siendo rigurosos, "The origin of the negative binomial distribution is not as a Poisson–gamma mixture, which is a rather new parameterization. The earliest definitions of the negative binomial are based on the binomial PDF. Specifically, the negative binomial distribution is characterized as the number of failures before the rth success in a series of independent Bernoulli trials. The Bernoulli distribution is, as you may recall, a binomial distribution with the binomial denominator set at one (1). Given r as an integer, this form of the distribution is also known as a Pascal distribution, after mathematician Blaise Pascal (1623–1662). However, for negative binomial models, r is taken as a real number greater than 0, although it is rarely above four." (Hilbe, 2011, pág. 3). Así, la primera NB en aparecer fue la NB2, aunque no exactamente como se le conoce hoy en día, aunque sí bajo la misma lógica: determinar el ***número de fracasos*** necesarios antes de una cantidad fija de éxitos.

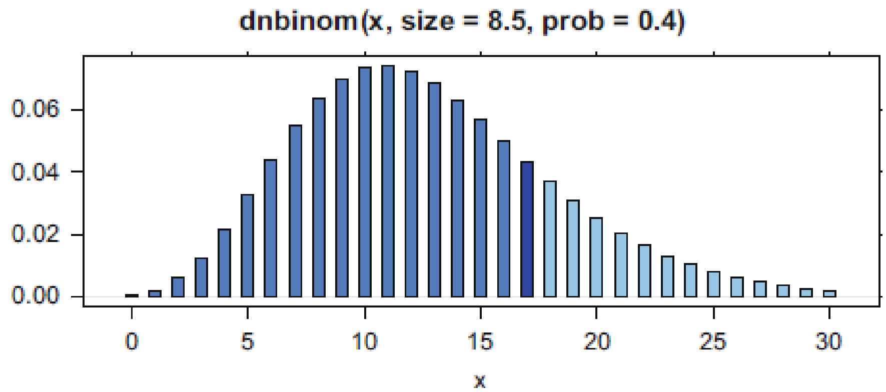
Por otro lado, la sintaxis en R correspondiente a la distribución binomial negativa 2 sería, según (Heiberger & Holland, 2015, pág. 825) y según las especificaciones matemáticas brindadas en el paquete informático, que pueden localizarse en (MIT, 2017), a la que se muestra en las figuras presentadas a continuación:

---

[39] En ambos autores la presentación inicial con la NB1 y tras su manipulación matemática se arriba a la NB2.

Figura 9

J.3.5 Negative Binomial




```
>   dnbinom(17, size=8.5, prob=.4)
[1] 0.04338

>   print(pnbinom(17, size=8.5, prob=.4), digits=17)
[1] 0.81209497223034977

>   qnbinom(0.81209497223034977, size=8.5, prob=.4)
[1] 17
```

Fuente: (Heiberger & Holland, 2015, pág. 825).

Figura 10

Description

Density, distribution function, quantile function and random generation for the negative binomial distribution with parameters `size` and `prob`.

Usage

```
dnbinom(x, size, prob, mu, log = FALSE)
pnbinom(q, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
qnbinom(p, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
rnbinom(n, size, prob, mu)
```

Fuente: (MIT, 2017).

Figura 11

Arguments

`x`	vector of (non-negative integer) quantiles.
`q`	vector of quantiles.
`p`	vector of probabilities.
`n`	number of observations. If `length(n) > 1`, the length is taken to be the number required.
`size`	target for number of successful trials, or dispersion parameter (the shape parameter of the gamma mixing distribution). Must be strictly positive, need not be integer.
`prob`	probability of success in each trial. `0 < prob <= 1`.
`mu`	alternative parametrization via mean: see 'Details'.
`log, log.p`	logical; if TRUE, probabilities p are given as log(p).
`lower.tail`	logical; if TRUE (default), probabilities are $P[X \le x]$, otherwise, $P[X > x]$.

Fuente: (MIT, 2017).

La sintaxis de la figura anterior exhibe la particularidad de utilizar una función de densidad y con ello estar habilitada para ingresar valores continuos en la función (otrora masa, ahora densidad), la explicación de ello obedece, según el contexto, al Teorema Central del Límite, a la Ley de los Grandes Números o al Teorema Ergódico; sin embargo, ello no pertenece estrictamente al foco de la investigación, por lo que se hablará sintéticamente sobre ello en los anexos.

Sobre la sintaxis de R en sí misma respecto a la NB2 el lector simplemente debe notar que para los fines de la presente investigación ya están definidas x, r y p (que en el paquete informático R aparece como $prob$), así como también r (que en el paquete informático R es $size$), mientras que respecto a la función cuantil y sobre la parametrización alternativa vía el primer momento de probabilidad se hablará en los anexos, por lo que no hay nada más que agregar al respecto.

IV.VI. Distribución NBII. Ejemplos Manuales y Automatizados con R Studio

Retomando el ejemplo localizado en (DeGroot & Schervish, 2012, pág. 297), (28) tomaría la siguiente apariencia:

$$f(x|4,p) = \begin{bmatrix} \left(\frac{(x+4-1)!}{(4-1)!\,(x)!}\right) p^4 (1-p)^x \end{bmatrix}, \qquad para\ todo\ x = 0,1,2,\dots,$$
$$0, \qquad en\ otro\ caso$$

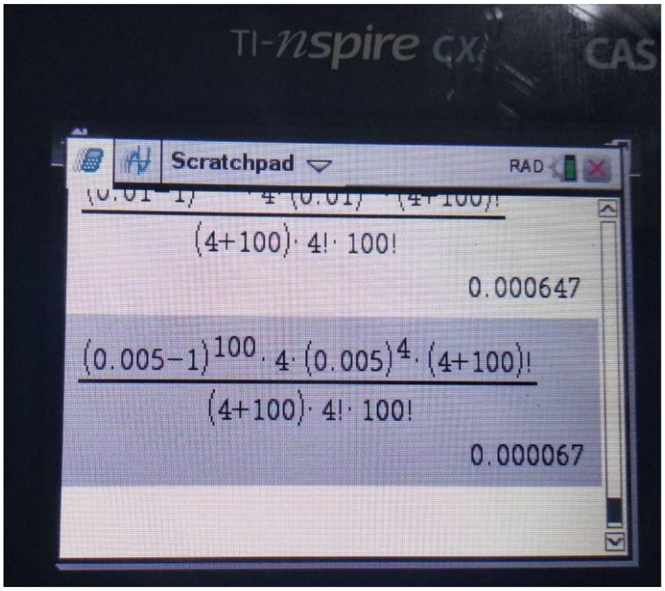
A continuación, mediante el uso de una Texas Instruments *n***spire** *CX* ***CAS***, sin recurrir a ninguna sintaxis particular de la misma, i.e., planteando las operaciones en términos de adiciones, diferencias, productos, cocientes, potencias y factoriales, se realizarán estimaciones para valores de x y de p seleccionados arbitrariamente con fines orientados puramente a la claridad expositiva.

El lector debe tener presente que cada ensayo de Bernoulli tiene una naturaleza intrínsecamente aleatoria (por ello poseen una familia de distribuciones de probabilidad que los describen) y, por consiguiente, deben tener una probabilidad individual de ocurrencia no relacionada (porque se asumen como independientes); tal probabilidad de ocurrencia se define en términos de dos conjuntos de ensayos. A tales conjuntos puede comprendérseles de la siguiente manera: se dispone del conjunto de ensayos que corresponden a los sucesos considerados como fracasos y del conjunto que lo complementa, es decir, del conjunto de ensayos que corresponde a los sucesos considerados como éxitos. Así, supóngase que cada ensayo de Bernoulli tiene una probabilidad de $p = 0.005$ y la cantidad de tales ensayos, es decir, el número de fracasos o eventos no deseados hasta la ocurrencia de 4 éxitos o eventos de interés (es decir, hasta que el sistema de alarmas reporte 4 veces que el capital fijo produjo una pieza defectuosa del bien de capital que produce) se estima de la siguiente forma:

$$f(100|4{,}0.005) = \left(\frac{(100+4-1)!}{(4-1)!\,(100)!}\right) 0.005^4 (1-0.005)^{100} = 0.000067.$$

Lo mismo puede verificarse en el instrumento especificado previamente:

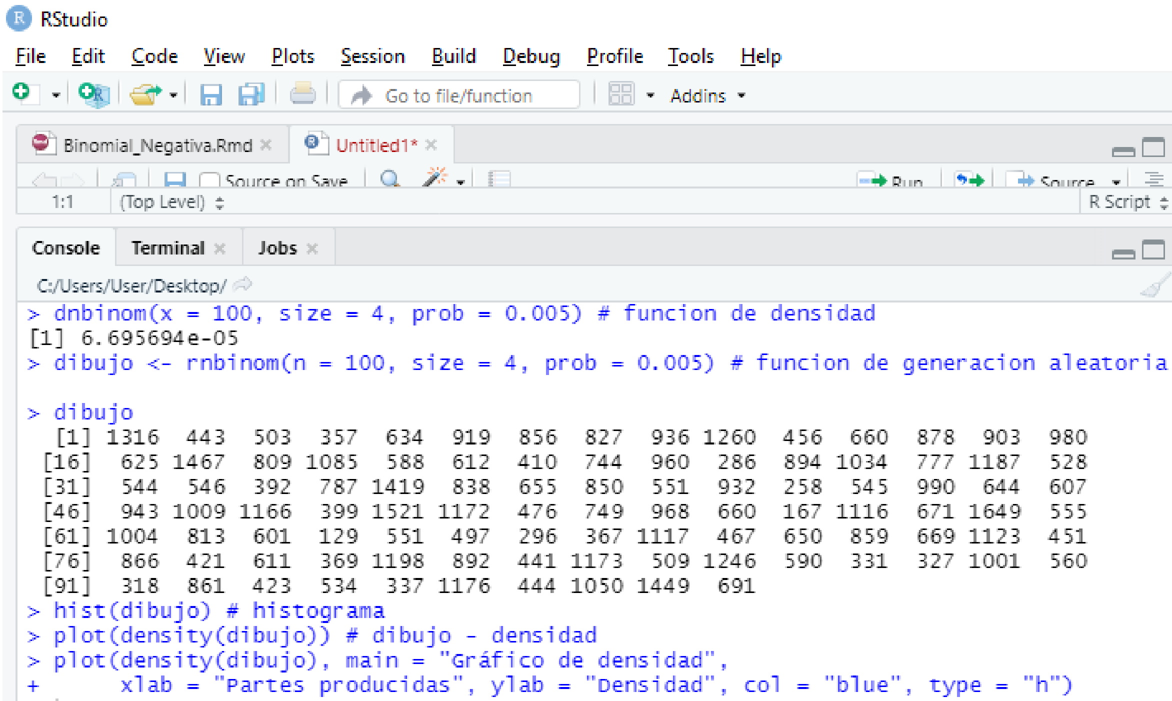
Figura 12



Fuente: Estimación propia mediante una Texas Instruments *n***spire** *CX* ***CAS***.

Por otro lado, la sintaxis en R para este ejercicio puede escribirse de la siguiente manera:

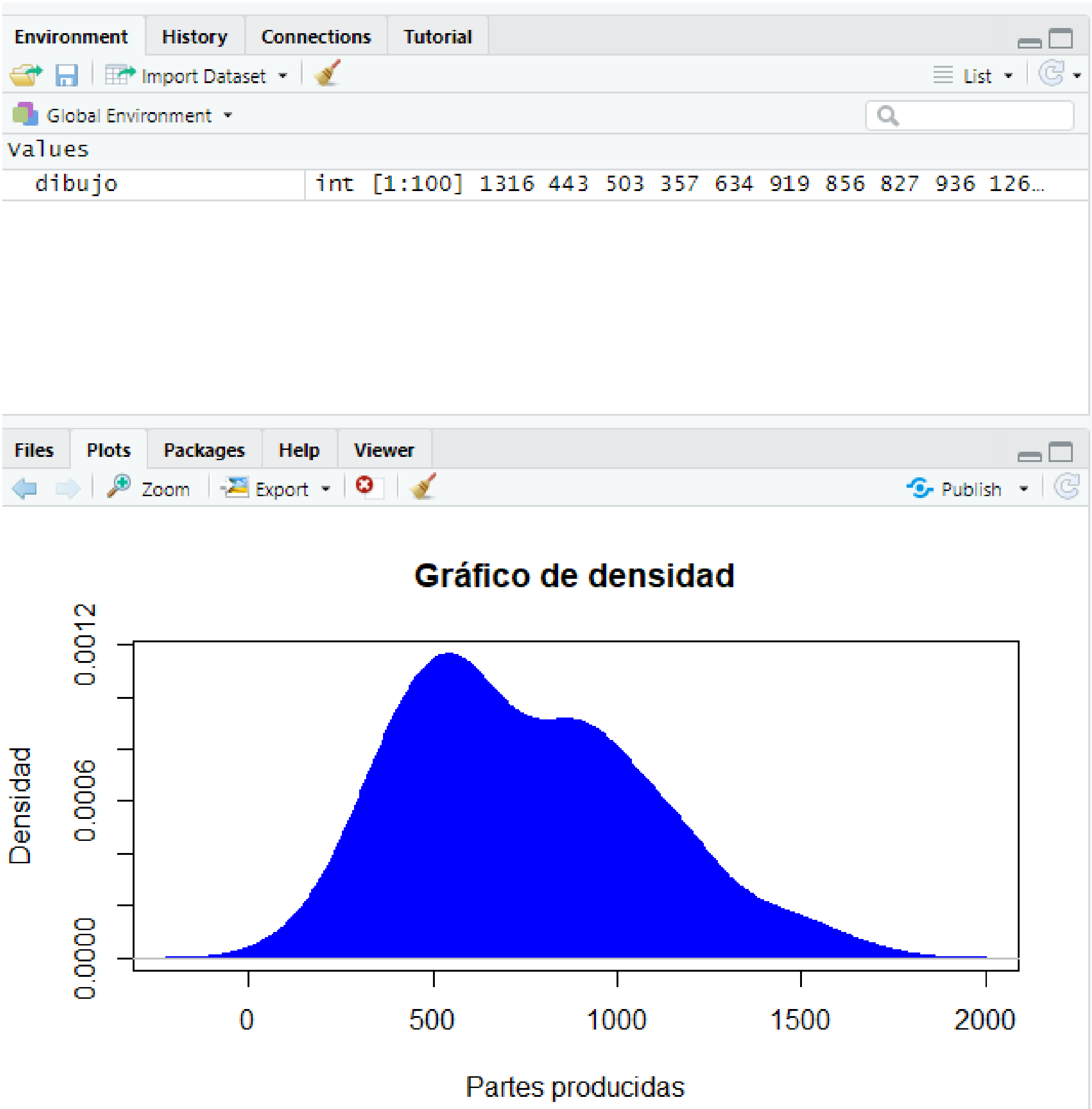
Figura 13



Fuente: Elaboración propia con ayuda de una colega que prefiere el anonimato.

Y los resultados de ella obtenidos pueden verificarse en la siguiente figura:

Figura 14



Fuente: Elaboración propia con ayuda de una colega que prefiere el anonimato.

Teniendo claro lo anterior y con la evidencia empírica disponible, la resolución del ejercicio anterior puede interpretarse como sigue:

"La probabilidad de que tras observar 100 eventos no deseados (conocidas sus probabilidades individuales de ocurrencia) el investigador observe 4 eventos de interés es de 0.000067."

Así, puede también construirse una forma de interpretación que generalice la anterior:

"La probabilidad de que tras observar $x \in X$ eventos no deseados (conocidas sus probabilidades individuales de ocurrencia p) el investigador observe r eventos de interés es de P".

Finalmente, es de importancia fundamental en la experimentación estadística el uso de sintaxis que permitan generar números pseudoaleatorios con la finalidad de simular en condiciones controladas el comportamiento de algún fenómeno natural y social, a través del uso de estas sintaxis para generar distribuciones de probabilidad, de forma tal que sea posible extraer de tales simulaciones información que contribuya a mejorar nuestro entendimiento de tales fenómenos de interés. En este sentido, con la ayuda del paquete informático R se plantea un ejemplo que permite mostrar ciertos aspectos eminentemente lógicos subyacentes a las distribuciones de probabilidad que únicamente se pueden observar cuando se replica de forma menos sofisticada su funcionamiento, así como también la forma en que los procesos son automatizados mediante los paquetes informáticos.

Para comprender mejor lo anterior, se explicará sintéticamente el Axioma de Elección con base en lo expuesto por (Billingsley, Probability and Measure, 1995, pág. 537). Anteriormente se dijo, con base en (Billingsley, Probability and Measure, 2012) en el lugar definido que el Axioma de Elección estaba implícito en los fundamentos de la Estadística Matemática y de la Teoría de Conjuntos, sin embargo, no se dieron detalles sobre este. Cualquiera que sea el conjunto de datos que un investigador esté analizando puede descomponerse como $[A_\theta : \theta \in \Theta]$, lo cual quiere decir que el conjunto de datos (expresado matemáticamente por el espacio muestral Ω) que estudia el investigador puede descomponerse[40] en

[40] Como se lee en (Daverman, 1986, pág. 1), "Technically speaking, the decompositions are all upper semicontinuous ones, meaning that the decomposition elements fit together in a fashion nice enough to ensure metrizability of the decomposition spaces.", en donde la metrizabilidad es simplemente la cualidad de un espacio (véase como un conjunto de objetos -para el caso de esta

diversos A_θ en donde el conjunto que los unifica se expresa mediante el símbolo Θ, que posee, entre otras características, que son todos los subconjuntos no vacíos que es posible formar con el espacio muestral Ω previamente definido. El Axioma de Elección de Zermelo es la hipótesis metamatemática[41] que sostiene que siempre es posible, sin importar el tamaño o cardinalidad del conjunto (pueden ser cardinales transfinitos, y ahí aparecen los problemas epistemológicos descritos anteriormente), elegir al menos 1 conjunto C que contenga exactamente 1 punto de cada A_θ: $C \cap A_\theta$ que es un conjunto unitario para cada θ formado de Θ. Lo anterior generaliza la intuición de siempre poder, sin importar el tamaño de un conjunto de datos, formar un conjunto de estos con al menos 1 elemento de cada uno de los conjuntos formados del conjunto original. Con base en ese principio se garantiza la

investigación tales objetos son datos-) de ser medible (establecer una distancia entre cualesquiera dos puntos mediante alguna función denominada *métrica* del espacio en cuestión). En libros de texto de Topología se encontrará esta temática bajo el nombre de "Decomposition of Manifolds" (como el título de la obra anteriormente citada" o como "Manifold Decomposition" y la razón de ello obedece a que *manifold* es la palabra en el idioma inglés utilizada para generalizar la palabra *surface*, dentro del mismo idioma, que es utilizada para hacer referencia a la generalización matemática del concepto superficie que de ordinario conocemos. De forma más intuitiva, recurriendo al concepto de esfera en la Geometría, "a manifold *M* may be decomposed or split by writing *M* as a combination of smaller pieces. When doing so, one must specify both what those pieces are and how they are put together to form *M*." (Wikipedia, 2020).

[41] Que surge para solventar la paradoja de Russell, que había aparecido anteriormente como la paradoja de Cantor, que es una forma más general de la primera (por involucrar directamente al conjunto universal y a su potencia, en donde la paradoja se genera porque la función sobreyectiva que se supone imposible encontrar del conjunto original a su potencia sí existe para este caso); la paradoja de Russell establece que el conjunto de conjuntos que no pertenecen a sí mismos puede pertenecer a sí mismo si y solo si no pertenece a sí, lo cual para casos particulares como la bolsa de bolsas que las personas suelen almacenar de sus visitas a los supermercados o para el caso de un conjunto de muñecas Matrioshka no habría problema en determinar la validez cognoscitiva de tal afirmación, sin embargo, se desdibuja para el caso de una bolsa de basura (¿reciclable o no? -y se pueden añadir más preguntas-) y definitivamente no aplica para el caso de una casa de campo que aloja personas o animales porque desdibujaría los estados de materia cualitativamente diferentes que existen de forma verificable en el mundo de la realidad. Lo anterior pone de manifiesto lo complejo del asunto y por qué es terreno de la Filosofía y no de las Ciencias, ni siquiera de las ciencias formales (la Teoría de Conjuntos no es Matemática Pura, es Metamatemática -que se define filosóficamente según la Escuela de Filosofía de las Matemáticas dominante, que tras la controversia entre Brouwer y Hilbert a inicios del siglo pasado en la que resultó "victorioso" el último, el Formalismo Matemático domina tal rama de la Filosofía de las Ciencias y de ello se explica que la Teoría de Conjuntos ZF-C haya progresado hasta dominar esta ciencia formal a pesar de que en su nacimiento generó rechazo por múltiples filósofos y polímatas, entre ellos Henry Poincaré-).

validez formal (que es parte de la validez epistemológica de un instrumento, más no la única y no necesariamente la más importante[42]) de todo el aparato teórico y aplicado de las Matemáticas.

---

[42] Existe toda una serie de lógicas y flexibilizaciones matemáticas de la versión formal de la Teoría de Conjuntos ZF-C que han demostrado ser robustas filosófica y matemáticamente (como el programa establecido por Quine), así como otras lógicas que flexibilizan el principio del tercero excluido (sobre eso trató la controversia Brouwer-Hilbert en los cimientos fundacionales de las Matemáticas) y permiten poder trabajar sin asumir este axioma. A nivel teórico se tiene el Análisis Constructivo de Bishop, a nivel aplicado todos los trabajos de medición realizados por los conjuntos difusos (*Fuzzy Sets*), conjuntos corrugados (*Rough Sets*) y los conjuntos fractales, que han dado origen a sus respectivas lógicas, así como también las investigaciones hechas desde la lógica cuántica (fundada por Neumann) y la lógica probabilística. Para ampliar más al respecto puede revisarse: a) https://www.anderstorvillbjorvand.com/_service/53/download/id/3378/name/19950428_project_report_fractal_logic.pdf, b) https://www.mathpages.com/home/kmath140/kmath140.htm, c) https://plato.stanford.edu/entries/qt-quantlog/, d) https://iep.utm.edu/qu-logic/, e) https://plato.stanford.edu/entries/logic-fuzzy/, así como también la entrada de *Lógica Probabilitaria* en (Rosental & Iudin, 1971). Respecto a las aplicaciones a nivel de aprendizaje automático (*Machine Learning*), aprendizaje profundo (*Deep Learning*) y minería de datos (*Data Mining*) puede (Bjorvand, 2020), localización que está descrita por su autor en los siguientes términos: "I did my master's degree in Computer Science in 1995/1996 at the Norwegian Institute of Technology (now: The Norwegian University of Science and Technology). From mid-1997 till mid-2000, I was working at the University of Oslo as a research scientist funded by The Research Council of Norway. In this time period, from 1995 till 2000, I published some articles at various international scientific conferences. Some of the papers have even been quite widely cited. So, if you are interested, you can download all of them here." Finalmente, si se desea conocer más sobre las aplicaciones de estos tipos de lógica, puede consultarse sus entradas en Wikipedia, aunque también existe literatura especializada al respecto, como por ejemplo (Smithson & Verkuilen, 2006) (que en la página 5 remite a Zimmerman (1993) y Klir & Yuan (1995)) e incluso un documento del descubridor de los conjuntos corrugados publicado en su época de investigador en el Departamento de Ingeniería Eléctrica y Ciencias de la Computación de la Universidad de Berkeley en 1975, se hace referencia aquí a (Zadeh, 1975). Además, para un estudio histórico de la Teoría de Conjuntos Corrugados y su lógica, puede consultarse en (Tamir, Rishe, & Kandel, 2015), cuya reseña podría dejárseles a los mismos autores en la página 3: "The concepts of truth and meaning are of fundamental importance in logic, information analysis and related fields. The theory outlined in this paper, call it RCT for short, is a departure from traditional theories of truth and meaning, principally correspondence theory, coherence theory, Tarski semantics, truth-conditional semantics and possible-world semantics (…) In large measure, traditional theories of truth and meaning are based on bivalent logic. RCT is based on fuzzy logic. Standing on the foundation of fuzzy logic, RCT acquires a capability to enter the realm of everyday reasoning and everyday dis-course a realm which is avoided by traditional theories of truth and meaning largely because it is a realm that does not lend itself to formalization in the classical tradition.In RCT, truth values are allowed to be described in natural language (…) Examples. Quite true, very true, almost true, probably true, possibly true, usually true, etc. Such truth values are referred to as linguistic truth values. Linguistic truth values are not allowed in traditional logical systems."

Lo anterior no es otra cosa que la garantía epistemológica[43] que las $permutaciones$, $combinaciones$, $k-permutaciones$ y $k-combinaciones$, así como los resultados de sus realizaciones, son congruentes con el cuerpo teórico existente de las Matemáticas sin fisuras (por supuesto, aceptando el supuesto *ad hoc* explicado en el párrafo anterior) y aunque puede ser un poco esquivo comprender la importancia de que un conocimiento sea válido epistemológicamente, generalizando a formas más flexibles (no lógico-formales) la validez epistemológica puede plantearse en términos negativos, para honrar el título de la investigación: ¿Acaso no es deseable poder estar seguros de que al realizar algo, lo que sea, una y otra vez sin importar las circunstancias, los resultados objetivos (pertenecientes a la realidad, los hechos en su verdad natural) obtenidos de realizar ese algo sean sistemáticamente los mismos? Esa es pues la utopía abstracta de la Filosofía Marxista y de las Ciencias.

Así, establecidos los cimientos teóricos-formales más profundos de las Probabilidades y la Teoría Estadística, se puede empezar a asociar a los conjuntos de diversas formas (ahí empieza a aparecer en toda su magnitud la importancia de la generalización de los instrumentos a utilizar) y de ello se desprenden las permutaciones y combinaciones, así como sus generalizaciones, lo cual como un todo constituye el conjunto de *Métodos de Conteo*, tal y como puede verificarse en lo expuesto por (Taboga, 2012, págs. 1-26).

En ese sentido se presenta a continuación un ejemplo teórico que compara el escenario de generar un conjunto de datos directamente a partir de la utilización

---

[43] "GNOSEOLOGÍA (…) o teoría del conocimiento. Parte importante de la teoría filosófica, versa acerca de la facultad del hombre para entrar en conocimiento de la realidad, acerca de sus fuentes, de las formas y de los métodos del conocimiento, acerca de la verdad y de los caminos para llegar a conocerla (…) la gnoseología materialista parte del reconocimiento del carácter objetivo del mundo exterior y de que es posible conocerlo (…) **(ello está ligado con)** el papel decisivo de la actividad de los hombres en la esfera de la producción social para el desarrollo del conocimiento (…) La dialéctica materialista, que descubre las leyes más generales del desarrollo de la Naturaleza, la sociedad y el pensar, nos ofrece la única teoría científica del conocimiento (…) Incluye en sí lo que en la actualidad se denomina teoría del conocimiento, gnoseología, lo cual también ha de examinar su objeto históricamente, estudiando y generalizando el origen y el desarrollo del conocer, el paso del no-saber al saber (…)" (Rosental & Iudin, 1971, pág. 204). El contenido en negrita ha sido adicionado por el autor de la presente investigación.

de métodos de conteo y otro conjunto de datos generarlo mediante las teorizaciones de las Probabilidades y la Estadística cristalizadas en las sintaxis de R diseñadas específicamente para la estimación relacionada con la distribución binomial negativa, las cuales han sido expuestas en pasajes anteriores de esta investigación.

Como se adelantó, el procedimiento presentado a continuación compara la generación de una distribución binomial negativa a partir de métodos de conteo cristalizados en las sintaxis personalizadas "experimento_CARTA" y "func_experimentos" con su estimación mediante las sintaxis automatizadas de R, con la finalidad de probar la equivalencia entre las sintaxis construidas vía método de conteo y las construidas vía Probabilidades y Teoría Estadística, demostrando su equivalencia y el valor de R (y de otros paquetes informáticos, evidentemente) como instrumento de medición y análisis. Evidentemente por motivos de estandarización, ambas metodologías se prueban bajo las mismas condiciones, fundamentalmente se deben mencionar: promedio, número de observaciones y cantidad de éxitos[44].

- **experimento_CARTA**: Esta función toma una carta de entre una baraja compuesta por 52 cartas (13 cartas de 4 palos) la cantidad de veces necesaria para que salga la carta indicada en el argumento de la función, el cual se llama "cartaBUSCADA". Por ejemplo, si "cartaBUSCADA" es igual a "3", entonces la función cuenta la cantidad de cartas que resulta necesario sacar para obtener un 3 (de cualquier palo).
- **func_experimentos**: Esta función ejecuta el experimento citado antes (experimento_CARTA) la cantidad de veces que se indica en el argumento, llamado "cantidad" y retorna un vector con los resultados del experimento.

[44] El investigador sólo explicado teóricamente el experimento, pero su diseño computacional se agradece a la colega antes citada, que insiste en permanecer en el anonimato.

Generación de Números Pseudo-Aleatorios mediante métodos de conteo que se distribuyen como una masa de probabilidad binomial negativa 2

```
cartas <- c("A", "2", "3", "4", "5", "6", "7", "8", "9", "10", "J", "Q", "K")
cartas52 <- rep(cartas, 4)
palos <- c("Picas", "Copas", "Diamantes", "Flores")
palos52 <- sort(rep(palos, 13))
baraja <- data.frame(cbind(cartas52, palos52))
carta <- unite(baraja, carta, sep = "_")
baraja <- cbind(baraja, carta)
escogencia <- unlist(carta)

experimento_CARTA <- function(cartaBUSCADA){
  cantidadCARTA <- 0
  final <- NULL
  cc <- 0
  while(cc == 0){
    c <- str_extract(sample(escogencia, size = 1), cartaBUSCADA)
    if(is.na(c)){
      cc = 0
    } else{
      cc = c
    }
    cantidadCARTA = cantidadCARTA + 1
    final <- c(final, cantidadCARTA)
  }
  return(final[length(final)])
}

experimento_CARTA("A") # ejemplo
```

```
## [1] 10
experimento_CARTA("2") # ejemplo
## [1] 9
func_experimentos <- function(cantidad){
 experimentos <- NULL
 for(i in 1:cantidad){
  experimentos <- c(experimentos, experimento_CARTA("A"))
 }
 return(sort(experimentos))
}

func_experimentos(100) # ejemplo
##   [1]  1  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  2  3  3  3  3
##  [26]  3  3  4  4  4  5  5  5  5  5  5  5  5  6  6  6  6  6  6  6  6  7  7  7  7
##  [51]  7  8  8  8  8  8  8  8  8  9 10 11 11 11 11 11 11 11 11 12 12 13 13 13 14
##  [76] 14 15 15 16 18 18 18 19 21 21 22 22 22 26 26 30 31 33 36 36 37 41 41 50 51
```

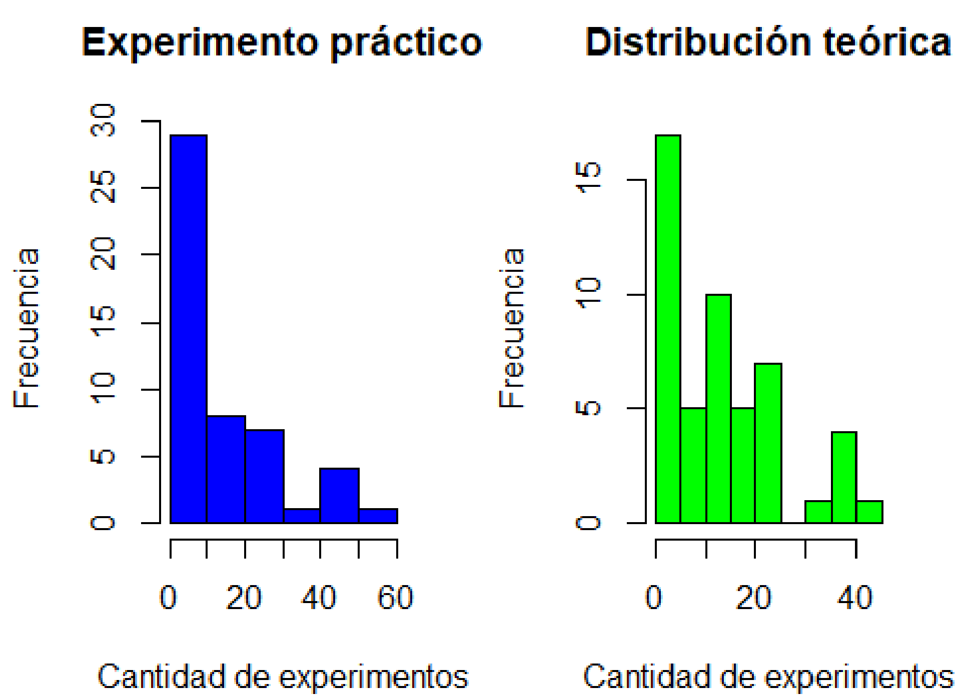
Generación de Números Pseudo-Aleatorios mediante transformaciones matemáticas de la función de distribución de la masa de probabilidad binomial negativa 2

Ahora se utiliza una función personalizada llamada **graficos**, la cual tiene como argumento "cantidad", el cual se refiere a cantidad de observaciones. Lo que esta función realiza es crear dos gráficos: Al lado izquierdo, aparece el histograma que resulta del ejercicio práctico realizado en el apartado anterior. Al lado derecho, se crea un histograma con la función **rnbinom**, que es la función de R que permite generar números aleatorios que siguen determinadas características. Los argumentos de rnbinom son aquellos que caracterizan a la distribución obtenida a partir del ejercicio práctico: mismo promedio, mismo número de observaciones y misma cantidad de éxitos. Los ejemplos muestran que los gráficos son muy similares, de modo tal que se confirma la teoría con la práctica.

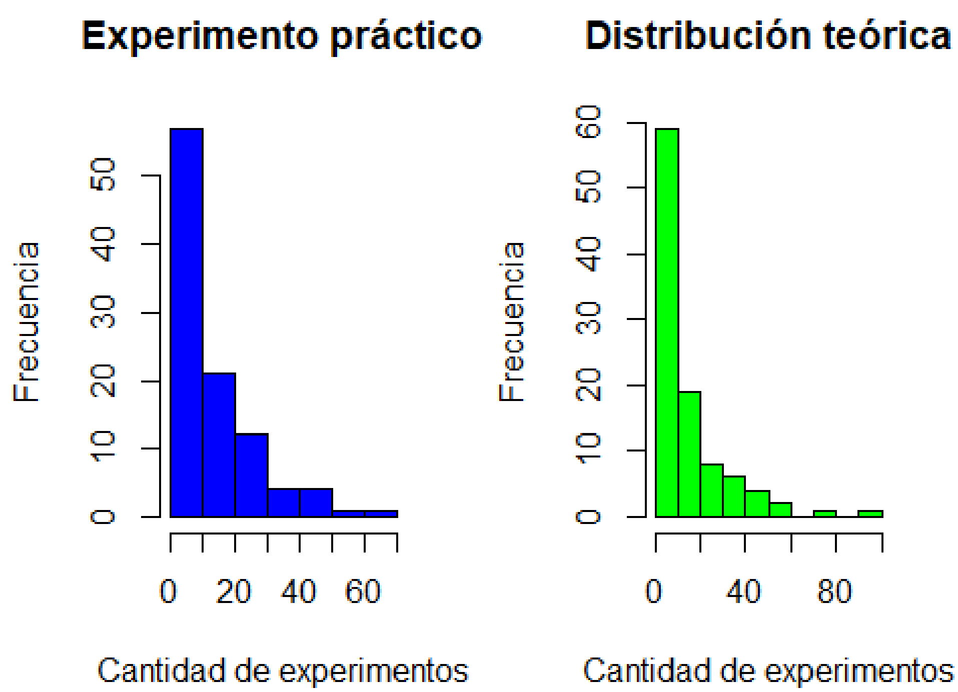
```
graficos <- function(cantidad){
 par(mfrow = c(1,2))
 grafico <- func_experimentos(cantidad)
 hist(grafico, col = "blue", main = "Experimento práctico",
     xlab = "Cantidad de experimentos", ylab = "Frecuencia")
 x1 <- rnbinom(cantidad, mu = mean(grafico), size = 1)
 hist(x1, col = "green", main = "Distribución teórica", xlab = "Cantidad de
experimentos", ylab = "Frecuencia")
}
```

Comparación Gráfica de ambos Generadores de Números Pseudo-Aleatorios

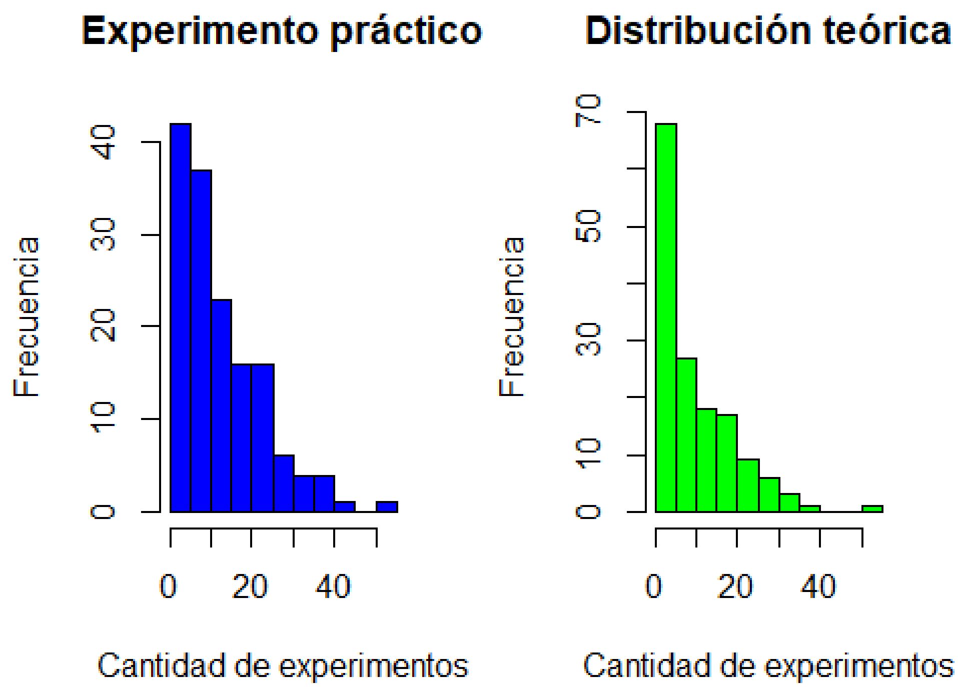
```
graficos(50) # ejemplo
```



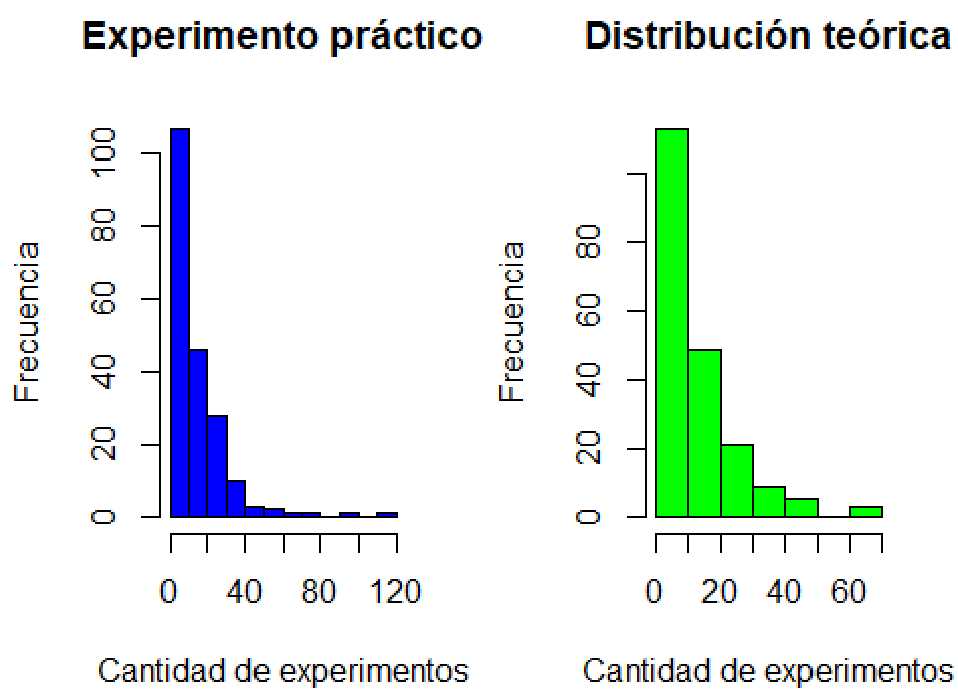
```r
graficos(100) # ejemplo
```




```r
graficos(150) # ejemplo
```




```r
graficos(200) # ejemplo
```



Finalmente, si el lector desea utilizar el programa informático *Python* y contrastar el mismo ejemplo hecho ahí (diferente al presentado aquí) con R, puede consultar (DexLab, 2020).

IV. ANEXOS

IV.I. Algunos Comentarios Sobre la Familia: Distribución Binomial Negativa

En la imagen presentada a continuación se muestran algunos de los miembros de esta populosa familia de distribuciones de masa de probabilidad.

IV.I. I. Algunas Relaciones Entre la Familia de Familias: Distribuciones Binomiales Negativas

Figura 15

5.1 Univariate distributions – tool coverage and code names

5.1.1 Tool coverage – discrete distributions

ProbOnto 2.5	Parame- -ters	UncertML 3.0	BUGS 1.4	Monolix 4.3	NONMEM 7.3	STAN 2.10.0
Negative Binomial 1	r, p	–	y	–	–	–
Negative Binomial 2	λ, τ	–	–	–	–	–
Negative Binomial 3	μ, ϕ	–	–	–	–	y
Negative Binomial 4	r, p	y	–	–	–	–
Negative Binomial 5	α, β	–	–	–	–	y
Negative Binomial 6	η, ϕ	–	–	–	–	y

- Relationship pair: $ConwayMaxwellPoisson1(\lambda, \nu) \rightarrow NegativeBinomial1(r, p)$
- Relationship type: Transformation
- Relationship definition: For $\nu = 0$ and $\lambda < 1$ the sum of Conway-Maxwell-Poisson distributed variables reduces to the sum of geometric variables, which follows a Negative Binomial distribution with parameters n and $1 - \lambda$

Relationships

- Relationship pair: $NegativeBinomial2(\lambda, \tau) \rightarrow NegativeBinomial5(\alpha, \beta)$
- Relationship type: Reparameterisation
- Relationship definition: $\alpha = 1/\tau, \beta = 1/(\tau\lambda)$
- Relationship reference(s): ProbOnto spec
- Relationship pair: $NegativeBinomial2(\lambda, \tau) \rightarrow NegativeBinomial4(r, p)$
- Relationship type: Reparameterisation
- Relationship definition: $r = 1/\tau, p = \frac{\tau\lambda}{1+\tau\lambda}$
- Relationship reference(s): ProbOnto spec [Cameron and Trivedi, 2013]
- Relationship pair: $NegativeBinomial2(\lambda, \tau) \rightarrow NegativeBinomial3(\mu, \phi)$

Fuente: Elaboración propia con base en (Swat, Grenon, & Wimalaratne, 2017, págs. 43, 86,197).

La figura anterior contiene únicamente 6 miembros de la familia (como se dijo en el cuerpo de esta investigación, algunos autores sostienen que son 13, otros que son más) y algunas de sus relaciones con los demás tipos de distribuciones. Esto puede ampliarse con mayor detalle en las fuentes referidas.

Finalmente, es fundamental expresar en un párrafo la relación de la familia de distribuciones binomial negativa 1 con la familia de densidades geométricas, así como su propiedad de *ausencia de memoria*, la cual es deseable en algunos campos de la investigación científica:

"The geometric family of densities is a subset of the family of negative binomial densities defined by setting $r = 1$. Thus, the geometric density is appropriate for assigning probability to events relating to how many Bernoulli trials are necessary to get the first outcome of type A. The geometric density function has a unique property in that it is the only discrete density for a nonnegative integer-valued random variable for which $P[(x > i + j | x > i)] = P[x > j]\ \forall i, j \in \{0,1,2, \dots\}$. This conditional probability property is referred to as the memoryless property, meaning that in any experiment characterized by the geometric density, if the experiment has already resulted in i trials without a type A outcome, the experiment has "no memory" of this fact, since the probability that more than j trials will be needed to obtain the first type A outcome is precisely the same as if the first i trials had never occurred. The proof that the geometric density has this property is left to the reader. The reader may wish to consult V.K. Rohatgi, (1976), *An Introduction to Probability Theory and Mathematical Statistics*, New York: John Wiley, p. 191, for a proof that the geometric density is the only density for nonnegative integer-valued random variables that has this property." (Mittelhammer, 2013, pág. 182).

IV.I. II. La Distribución Binomial Negativa I

Como se vio anteriormente, la NB2 es la forma funcional de la familia de distribuciones binomiales negativas que permite analizarse en términos de modelos jerárquicos de tipo mixtura de distribuciones de probabilidad, que fue el enfoque central de esta investigación, que es una relación dialéctica mediata.

Bajo este enfoque no quedó explícita, completamente tangible, su vinculación con la distribución binomial positiva, precisamente porque era un análisis mediato[45]. La distribución binomial negativa 1 posee relaciones inmediatas con la distribución binomial positiva y por ello puede facilitar la comprensión de los vínculos entre ambas familias de distribuciones de probabilidad.

Según (Mittelhammer, 2013, pág. 181), lo que puede corroborarse en (Wackerly, Mendelhall III, & Scheaffer, 2010, págs. 121-122) con otra simbología y con el coeficiente binomial que le corresponde en su forma compacta, la forma funcional de la NB1 se define de la siguiente manera:

$$f(x|r,p) = \begin{matrix} \left[\left(\frac{(x-1)!}{(r-1)!\,(x-r)!}\right)p^r(1-p)^{x-r}\right], \quad para\ todo\ x = r, r+1, r+2, \dots, \\ 0, \quad en\ otro\ caso \end{matrix} \tag{30}$$

Así, (30) es la forma funcional de la NB1 y los parámetros individualmente analizados poseen el mismo significado que para el caso de la NB2. El significado de que x se defina en términos de r se explicará durante la exposición de un caso aplicado simple.

La distribución NB1 posee una relación dialéctica inmediata con la distribución binomial positiva, por lo que sus vínculos, aunque no más profundos, pueden resultar más visibles para tal o cual investigador.

Como se señala en (Mittelhammer, 2013), la NB1 se puede utilizar para construir un espacio de probabilidad diseñado en vistas de un experimento que consta de $n-$ é$simas$ repeticiones independientes de un experimento dado del tipo de Bernoulli (ensayos de Bernoulli) al igual que en el caso de la densidad binomial positiva 1, excepto que la información que se desea conocer ahora es cuántos ensayos de Bernoulli son necesarios para obtener r resultados de un tipo

[45] Este es el nombre dado a la distribución binomial ordinaria en (Greenwood & Yule, 1920, pág. 274).

particular, digamos, del tipo A (por ejemplo, ¿cuántos ensayos de Bernoulli son necesarios para obtener r éxitos, fallos o colas?). Al comparar la densidad binomial positiva 1 con la densidad binomial negativa 1, ***observe que los roles del número de ensayos de Bernoulli y el número de éxitos se invierten con respecto a cuál es la variable aleatoria y cuál es el parámetro. Para la densidad binomial negativa 1, el valor de $\boldsymbol{x}$ representa el número de ensayos de Bernoulli necesarios para obtener $\mathbf{r}$ resultados de tipo A***[46, 47], mientras que para la "densidad" binomial negativa 2 la definición de x se restringe para el número de eventos/ocurrencias no deseadas o fracasos. Así, queda en evidencia su nexo con la distribución binomial positiva[48].

Un Simple Caso de Aplicación[49]

Un estudio geológico indica que un pozo petrolero de exploración perforado en una región particular debe producir petróleo con probabilidad de 0.2. El investigador desea encontrar la probabilidad de que el tercer descubrimiento de petróleo llegue en el quinto pozo perforado.

Suponiendo que las perforaciones independientes y la probabilidad de éxito en cada una de ellas es $p = 0.2$, es decir, de descubrir petróleo en cualquiera de los pozos, puede denotarse con Y el número del intento en el que ocurre el tercer descubrimiento de petróleo. Además, supóngase que se ha hecho el estudio necesario y suficiente que verifica que el fenómeno estudiado en su versión probabilística se comporta de forma binomial negativa 1 con $p = 0.2$, $r = 3$ y $y = 5$.

---

[46] En donde A no es una variable ni un parámetro, sino la cualidad o conjunto de cualidades de los $r - \acute{e}simos$ resultados que los hacen deseables (ser considerados como éxitos) para el investigador.
[47] La lógica bajo la cual Mittelhammer llama densidades a las masas de probabilidad consiste en concebir a las masas de probabilidad en una forma más general, i.e., viéndolas como si fuesen densidades de probabilidad con características particulares que las hacen discontinuas en determinados puntos. Por supuesto, lo ideal es mantener la distinción entre masas y densidades.
[48] Como el lector puede verificar en (Swat, Grenon, & Wimalaratne, 2017, pág. 43), siendo rigurosamente técnicos y retomando (Greenwood & Yule, 1920, pág. 274), la distribución binomial comúnmente conocida no debería llamarse solamente distribución binomial positiva, sino *distribución binomial positiva 1*.
[49] Tomado de (Wackerly, Mendelhall III, & Scheaffer, 2010, págs. 122-123). Ahí el lector notará que en lugar de $1 - p$ aparece q, sin embargo, son perfectamente equivalentes.

Dado lo anterior, es posible escribir (30) sustituyendo los valores definidos:

$$f(y|r,p) = f(5|3,0.2) = \left[\left(\frac{(5-1)!}{(3-1)!\,(5-3)!}\right)(0.2)^2(0.8)^{5-3}\right] = 0.307$$

Si se desea verificar el procedimiento anterior en paquete estadístico R, el lector puede utilizar la misma sintaxis que para el caso de la NB2 para el caso de la NB2, sustituyendo y por $y_0 - r$, es decir, agregando simplemente la siguiente modificación: $dnbinom(y_0 - r, 3, 0.2)$, lo cual obedece a que en (30) se definió en términos de $y = r, r+1, r+2, \ldots$,. Para lo referente a mayores niveles de caracterización de la distribución, por ejemplo, la estimación de sus momentos, el lector puede consultar la bibliografía de esta investigación.

Finalmente, en la figura presentada a continuación se resume lo planteado en el párrafo anterior:

Figura 16

Binomial:

- Fixed number of trials (*n*)
- Fixed probability of success (*p*)
- Random variable is X = Number of successes.
- Possible values are 0 ≤ *X* ≤ *n*

Negative Binomial:

- Fixed number of successes (*r*)
- Fixed probability of success (*p*)
- Random variable is Y = Number of trials until the *r* th success.
- Possible values are *r* ≤ *Y*

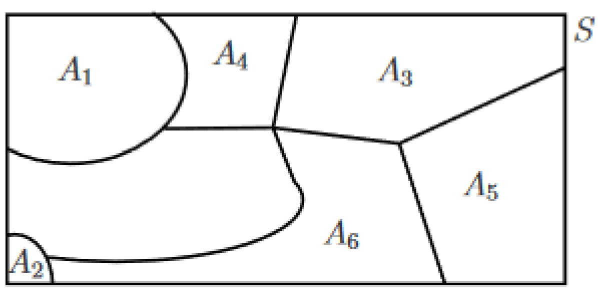
Fuente: (Stack Exchange, 2020).

IV.II. Algunos Comentarios Sobre Probabilidad Total, Probabilidad Inversa y Bayesianismo Objetivo

IV.II. I. Probabilidad Total

Como puede verificarse en (Loughborough University, 2008, pág. 2), la probabilidad total consiste, en esencia, en “a partition is a collection of non-empty, non-overlapping subsets of a sample space whose union is the sample space itself.”[50] En lugar de una demostración formal, se expondrá la intuición geométrica que expresa esa demostración en las figuras presentadas a continuación:

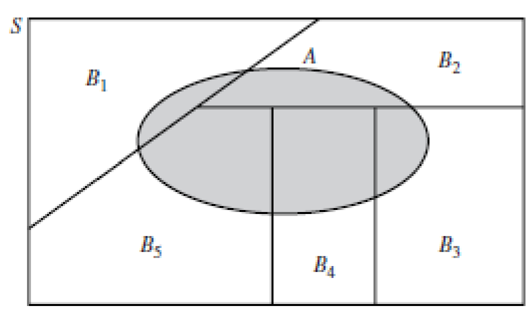
Figura 17



Fuente: (Loughborough University, 2008, pág. 2).

Figura 18

Figure 2.2 The intersections of A with events $B_1, \ldots, B_5$ of a partition in the proof of Theorem 2.1.4.



Fuente: (DeGroot & Schervish, 2012, pág. 60)

---

[50] Las diferentes notaciones en que esto puede expresarse se localizan tanto en la fuente en cuestión, como en (DeGroot & Schervish, 2012, pág. 60), (Mittelhammer, 2013, pág. 30), (Feller, 1968, pág. 22), aunque en este último se encuentra en los mismos términos que en (Loughborough University, 2008, pág. 2), en un sentido más topológico (aunque con mayor rigurosidad lógico-formal que en la universidad citada).

Del teorema de la probabilidad total se desprende la formulación no bayesiana de la probabilidad condicional:

$$\Pr(A|B) = \frac{\Pr(\mathrm{A} \cap B)}{\Pr(B)}, siempre\ que\ P(B) > 0 \qquad (31)$$

La identidad (31) expresa la probabilidad condicional de A dado B, mientras que la siguiente identidad expresa la probabilidad condicional de B dado A:

$$\Pr(B|A) = \frac{\Pr(\mathrm{B} \cap A)}{\Pr(B)} = \frac{\Pr(\mathrm{B} \cap A)}{\Pr(A)}, siempre\ que\ P(A) > 0 \qquad (32)$$

Por lo anteriormente realizado, puede verse que la aritmética es rigurosamente clara respecto a la equivalencia de los numeradores en (31), por lo que es posible re-escribir (30) sustituyendo su numerador por el numerador $\Pr(\mathrm{B} \cap A)$ de (31), conociendo de antemano su rigurosa validez matemática por la misma demostración lógico-formal del teorema localizada en (DeGroot & Schervish, 2012, pág. 77). Con ello, se arribaría a:

$$\Pr(B|A) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}, \qquad siempre\ que\ \Pr(A), \Pr(B) > 0\ , \qquad (33)$$

El lector debe notar que (33) es equivalente a (7). La novedad del teorema de Bayes es que conecta explícitamente las probabilidades de ocurrencia de dos fenómenos diferentes y por ello ahora no sólo el denominador debe ser estrictamente positivo, sino también $\Pr(A)$. El lector seguramente se preguntará en términos intuitivos ¿acaso no siempre fue así?, la respuesta es sí y no. Intuitivamente hablando sí, sin embargo, aunque en el diseño de un instrumento matemático la intuición juega el papel fundamental (es la inspiración, fuente de construcción conceptual del instrumento, etc.), al momento de la construcción del instrumento como tal esta pasa a ser sometida a la Lógica Formal, que entre más se generaliza a sí misma y en sí misma (como sistema axiomático y en sus construcciones individuales) suprime progresivamente las intuiciones que la generaron[51], lo que ocasiona que los vínculos de la realidad con el instrumento que de su estudio (el de la realidad) se generó y cuyo carácter no es inmediato[52] se desdibujen formalmente y es ahí donde entra a jugar su papel clásico la Filosofía, aquí se hace referencia a la Filosofía Marxista, la única filosofía científica que existe[53]. Sin embargo, al desdibujarse tales intuiciones, también puede ocurrir que se modifique la capacidad del instrumento (ahora generalizado) para extraer de su uso las conclusiones originalmente deseadas[54] o puede ocurrir que inmediatamente se formalice una intuición esta se

[51] Aquí ocurren diferentes y distintas cuestiones que no se tratarán a profundidad porque en sí mismas exigirían una investigación orientada a ellas (mucho más profundas que esta) y se desdibujaría la línea entre la Filosofía y la Filosofía de la Estadística; justificar ello en contexto de la presente investigación demanda más recursos que los actualmente disponibles. Sin embargo, el núcleo de esta cuestión puede encontrarse estudiando los siguientes momentos históricos: 1) la polémica entorno a la teoría de conjuntos de Cantor cuando apareció, 2) La aparición del Axioma de Elección de Zermelo y la reacción de los matemáticos y filósofos de la época, 3) la Controversia Brouwer-Hilbert, 4) la obra de Errett Bishop, 5) la reseña de una obra de cálculo infinitesimal de espíritu filosófico formalista que Paul Halmos le encomendó a Bishop.

[52] Por ejemplo, la intuición geométrica de espacios de dimensiones infinitas se encuentra en los espacios euclidianos publicados en la obra de Euclides en la época de la antigua Grecia, pero formalmente esas intuiciones han desaparecido si se habla, por ejemplo, de funtores; o bien, si se habla de espacios en que la métrica no se induce por una norma sino al contrario, etc. Sin embargo, la intuición se encuentra ahí, esfuerzos como la presente investigación buscan llevarla a la superficie.

[53] "(…) yo no creo en la verdad absoluta del Marxismo en el sentido en que algunas personas creen en los dogmas religiosos. Yo sólo creo que está lo suficientemente cerca de la verdad para hacerlo digno de apostar mi vida a él en contra de las teorías rivales." (Haldane, 1945, pág. 257) dijo el biólogo marxista hindú que restableciera la selección natural como el mecanismo esencial del cambio evolutivo (explicándolo en términos de las consecuencias matemáticas de la genética mendeliana) nacido en el día de Guy Fawkes en 1892; la traducción ha sido realizada por el autor de esta investigación y los datos biográficos fueron tomados de (Wikipedia, 2020).

[54] Un ejemplo extremo de esto puede verse cuando Jerzey Neyman y Egon Pearson investigaron en una forma de generalizar la prueba de hipótesis propuesta por Ronald Fisher, como se verifica en (Perezgonzalez, 2015, pág. 3). De hecho, es en ese contexto en que aparece con más fuerza la necesidad redefinir las creencias racionales de forma más intuitiva. Es una creencia racional aquella formada por el ser humano en su faceta investigador (aunque no necesariamente estarán respaldadas desde el oficialismo científico) cuyo fundamento sea la práctica científica y que la lógica bajo la cual procese su práctica científica para transformarla en experiencia, en *criterio experto*, sea una lógica fundamentada en estudios de planteamientos del marco teórico de la ciencia en cuestión y de las prácticas científicas suscitadas al interior de ella. El criterio experto y la información histórica son elementos fundamentales en el *meta-análisis* planteado por Ronald Fisher, que es ir mucho más allá de la prueba de hipótesis estándar, como se señala en. Esto va acorde también a lo planteado por (Gigerenzer, 2004, pág. 599), quien al respecto dice "At issue here is the importance of good descriptive and exploratory statistics rather than mechanical hypothesis testing with yes–no answers. Good descriptive statistics (as opposed to figures without error bars, or

desdibuje en su conexión con la realidad que la originó, como parece ser el caso de las probabilidades condicionales previo al teorema de Bayes. Fue hasta Bayes que esa intuición se reflejó en el instrumento con la suficiente claridad para explicitar las conclusiones que en la actualidad se extraen del mismo, fundamentalmente en el contexto de prueba de hipótesis, que como señala (Russell, 2014) es el interés del día a día en la praxis científica[55]. Son estas conclusiones las que han dado origen a la cuantiosa y altamente sofisticada familia de métodos estadísticos conocida como *Estadística Bayesiana*.

IV.II. II. Probabilidad Inversa y Bayesianismo Objetivo

En su génesis histórica, no solo el teorema de Bayes sino también la totalidad de la obra de Bayes sobre probabilidades nace para abordar lo que se conoce *probabilidad inversa*. Como se señala en (Wikipedia, 2020), el término probabilidad inversa es el término designado para la distribución de probabilidad de una variable no observada. En la actualidad, al hecho de determinar probabilísticamente una variable no observada (en un experimento sobre alguna variable observada) se le llama *probabilidad bayesiana*, en donde la "distribución"[56] de probabilidad de los datos (o variable aleatoria observada), dada la variable aleatoria no observada, es (siendo rigurosamente técnicos) la función de verosimilitud[57], mientras que la distribución de probabilidad de la variable no observada es, dado el conjunto de datos y la distribución a priori o prior, la distribución posterior contenida en el teorema de Bayes visto en secciones anteriores. En suma, lo que antes era conocido como probabilidad inversa ahora es lo que se expuso en las secciones previas y se etiquetó en ellas como probabilidad condicional en su versión bayesiana,

---

unclear error bars, and routine aggregate instead of individual analysis, for example) is necessary and mostly sufficient."

[55] Entiéndase esta como la comunión de las teorías científicas con las prácticas científicas.
[56] Las comillas obedecen a que no es una distribución como tal, como se verá a continuación.
[57] Que no es una distribución de probabilidad sino una transformación realizada sobre ella para optimizarla, en donde la optimización se suscita, para este caso, como una maximización. Lo anterior es la lógica fundamental de la familia de métodos que estiman las probabilidades por máxima verosimilitud y sus derivados mediatos e inmediatos.

denominada también como *probabilidad bayesiana*, que tiene tanto una interpretación objetiva como una subjetiva, sobre lo que hay que decir algunas cuestiones.

La definición de un grado racional de creencia debe entenderse como se señala a continuación: "The *objective Bayesian interpretation of probability* is the construal of probabilities in terms of rational degrees of belief, where these rational degrees of belief are given an objective Bayesian account (…) In sum, rational degree of belief is relative to evidence and language. Degrees of belief are rational if they are fit-for-purpose. Plausibly, degrees of belief are rational if and only if they are determined in the right way from evidence and language, and if in the first place the agent is rational to grant that evidence and adopt that language (in so far as the agent has any choice about her evidence and language). Objective Bayesian epistemology concerns only the former question: the link between evidence and language on the one hand, and rational degree of belief on the other; it does not concern choice of evidence or choice of language. Objective Bayesianism is a theory which holds that the degrees of belief which best fit their purpose are those that are probabilistic, calibrated with evidence, and otherwise equivocal." (Williamson, 2010, págs. 9-10).

Además, las características cualitativas de la probabilidad bayesiana objetiva (sujetas a la definición general de probabilidad establecida al comienzo de esta investigación) son retomando lo planteado en (Williamson, 2010, págs. 11-12):

1) Objetividad en su sentido filosófico. "We come to know about probabilities in various ways: we measure population frequencies, we appeal to symmetry arguments or scientific theories, we make educated guesses, and we derive some probabilities from others using the probability calculus. A philosophical theory of probability should explain how we can use such techniques to discover probabilities. If the theory rejects some of these

techniques, it should say where they go wrong and why they are apparently successful."

2) Objetividad en su sentido lógico. "For example, the probability that a patient's breast cancer will recur after treatment apparently depends on features of the cancer, of the treatment, and of the patient. It is not simply a matter of personal opinion: if two prognostic probabilities differ, at least one of them must be wrong. A philosophical interpretation of probability should, if possible, yield a notion of probability that is suitably objective in this logical sense—otherwise, it is revising rather than faithfully interpreting probabilistic statements as they occur in these applications."
3) Computabilidad en relación con su interpretación filosófica. "Probabilities are manipulated and inferences are drawn from them by means of the probability calculus. This mathematical apparatus, based on axioms put forward by Kolmogorov (1933) has by now become well entrenched. Consequently, a philosophical interpretation of probability should yield a notion that satisfies the axioms of probability. Otherwise, it is not a theory of probability—it is a theory of something else." Por supuesto, el lector no debe entender que "yield" implica una identidad, puesto esto implicaría previamente la identidad entre la Filosofía y las Matemáticas, lo cual es evidentemente falso y afortunadamente imposible. Aquí se habla en el sentido en que (Feller, 1968, pág. 1) plantea de que las intuiciones deben estar en unidad y relación con la parte lógico formal y con sus aplicaciones (en referencia a la evidencia empírica), aunque evidentemente esta unidad no es posible que sea lógico-formal y eso explica en alguna medida el porqué, como el mismo Feller señala en las primeras 3 líneas del prefacio a la tercera edición, es decir, en (Feller, 1968, pág. vii), de que fuera de la Unión Soviética muy pocos matemáticos tuviesen un buen concepto de las Probabilidades como una rama legítima de las Matemáticas.

4) Capacidad de Interpretación General. "A philosophical theory of probability should be able to cope with this variety—it should account for each use of probability, or, if some uses are to be viewed as illegitimate, it should say how such uses should be eliminated in favour of the legitimate uses. Otherwise, the theory is at best a partial theory, a theory of some of the uses of probability."

Para profundizar sobre las variantes epistemológicas del Bayesianismo subjetivo, puede consultarse (Williamson, 2010, pág. 15). Finalmente, si se desea ver una forma compacta de la forma funcional de las sumas e integrales de las probabilidades condicionales estudiadas en esta investigación (con la notación estándar de los cursos de Cálculo universitarios), el lector puede consultar (Greene, 2012, pág. 1070).

IV.II. III. Sobre los Momentos Muestrales de la Distribución Binomial Negativa II

Como se observa en (Cohen & Cohen, 2008, pág. 166), para la estimación de los primeros dos momentos muestrales de la NB2 se debe expresar esta de forma más general:

$$P(X = x) = \left(1 + \frac{m}{k}\right)^{-k} \frac{\Gamma(k + x)}{x!\,\Gamma(k)} \left(\frac{m}{m + k}\right)^{x} \tag{34}$$

En la identidad (34), $\Gamma(\alpha)$ es la denominada *función gamma*. Esta función es definida mediante:

$$\int_0^{\infty} e^{-x} x^{\alpha-1} dx$$

Mientras que para el caso discreto se define mediante:

$$\Gamma(n) = (n-1)!$$

Así, es posible estimar los momentos muestrales estimando m y k, puesto que:

Primer Momento Muestral (Nombre: Media Aritmética Muestral)

$$\widehat{m} = \bar{X}$$

Segundo Momento Muestral (Nombre: Varianza Muestral)

$$\hat{k} = \frac{\bar{X}}{S^2 - \bar{X}}$$

V. REFERENCIAS

Agarwal, A., Bajorski, P., Farnsworth, D. L., Marengo, J. E., & Qian, W. (2017). The Conditional Poisson Process and Negative Binomial Distributions. *Open Journal of Statistics, 7*, 16-22.

Bagui, S., & Mehra, K. L. (2019). On the Convergence of Negative Binomial Distribution. *American Journal of Mathematics and Statistics, 9*(1), 44-50. doi:10.5923/j.ajms.20190901.06

Bayes, T. (1763, Diciembre 23). An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, 370-418.

Bernoulli, J. (2006). *The Art of Conjecturing (Together to a Friend on Sets in Court Tennis).* Maryland: John Hopkins University Press.

Billingsley, P. (1995). *Probability and Measure.* New York: John Wiley & SONS.

Billingsley, P. (2012). *Probability and Measure* (Anniversary Edition ed.). John Wiley & Sons, Inc., Publication.

Bjorvand, A. T. (2020, Septiembre 29). *Anders Torvill Bjorvand*. Retrieved from My research: https://www.anderstorvillbjorvand.com/myresearch

Casella, G., & Berger, R. L. (2002). *Statistical Inference* (Segunda ed.). Pacific Grove: Thomson Learning.

Cheung, M. W.-L. (2015). *Meta-Analysis. A Structural Equation Modeling Approach.* West Sussex: John Wiley & Sons, Ltd.

Cohen, Y., & Cohen, J. Y. (2008). *Statistics and Data with R (An Applied Approach Through Examples.* West Sussex, United Kingdom: John Wiley & Sons Ltd.

Cook, J. (2009, October 28). *Negative Binomial.* Retrieved from Notes on the Negative Binomial Distribution: https://www.johndcook.com/negative_binomial.pdf

Daverman, R. J. (1986). *Decompositions of Manifolds.* Orlando: ACADEMIC PRESS, INC.

de Moivre, A. (1718). *The Doctrine of Chances or, A Method of Calculating the Probability of Events in Play.* London: Printed by W. Pearfor, for the Author. MDCCXVIII. Retrieved from https://books.google.it/books?id=3EPac6QpbuMC&printsec=frontcover#v=onepage&q&f=false

DeGroot, M., & Schervish, M. (2012). *Probability and Statistics.* Boston: Pearson Education.

DexLab. (2020, Septiembre 30). *Statistical Application in R & Python*. Retrieved from Negative Binomial Distribution: https://www.dexlabanalytics.com/blog/statistical-application-in-r-python-negative-binomial-distribution

Efron, B. (1978). Controversies in the Foundations of Statistics. *The American Mathematical Monthly*, 231-246.

Encyclopaedia Brittanica. (2020, Septiembre 27). *Science - Physics - Physicist*. Retrieved from Siméon-Denis Poisson: https://www.britannica.com/biography/Simeon-Denis-Poisson

Eremenko, A. (2020, Abril 30). *Stack Exchange, History of Sciences and Mathematics*. Retrieved from What was Kolmogorov's point of view in the philosophy of mathematics?: https://hsm.stackexchange.com/questions/11730/what-was-kolmogorov-s-point-of-view-in-the-philosophy-of-mathematics

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (Tercera ed., Vol. I). New York: John Wiley & Sons, Inc.

Gigerenzer, G. (2004). Mindless Statistics. *The Journal of Socio-Economics*, 587-606.

Greene, W. H. (2012). *Econometric Analysis (International Edition).* Essex: Pearson Education Limited.

Greenwood, M., & Yule, G. U. (1920, Marzo). An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society, LXXXIII*(2), 255-279. Retrieved from https://www.jstor.org/stable/2341080

Haldane, J. B. (1945). *Science and Everyday Life.* Allahabad,: Kitab Mahal Publishers.

Hegel, F. (1968). *Ciencia de la Lógica.* Buenos Aires: Solar / Hachette.

Heiberger, R. M., & Holland, B. (2015). *Statistical Analysis and Data Display (An Intermediate Course with Examples in R)* (Segunda ed.). New York: Springer.

Hilbe, J. M. (2011). *Negative Binomial Regression* (Segunda ed.). Cambridge: Cambridge University Press.

Hilfer, R. (2010, Julio 2). *Universidad de Pau y Pays de l'Adour.* Retrieved from Threefold Introduction to Fractional Derivates: http://lma.univ-pau.fr/meet/dfa2010/slides_hilfer.pdf

Ibe, O. C. (2013). *Markov Processes for Stochastic Modeling* (Segunda ed.). California: Elsevier.

Joyce, D. (2014, Agosto 12). *Clark University, Department of Mathematics and Computer Science*. Retrieved from Euclid's Elements. Book IX. Proposition 35: https://mathcs.clarku.edu/~djoyce/java/elements/bookIX/propIX35.html

Keith, T. Z. (2019). *Multiple Regression and Beyonds. An Introduction to Multiple Regression and Structural Equation Modeling* (Tercera ed.). New York: Taylor & Francis.

Kim, A. (2019, Junio 1). *Towards Data Science*. Retrieved from Poisson Distribution — Intuition, Examples, and Derivation: https://towardsdatascience.com/poisson-distribution-intuition-and-derivation-1059aeab90d

Kim, A. (2019, Agosto 6). *Towards Data Science*. Retrieved from Exponential Distribution — Intuition, Derivation, and Applications: https://towardsdatascience.com/what-is-exponential-distribution-7bdd08590e2a

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (Cuarta ed.). New York: The Guildford Press.

Kolmogórov, A. (1956). *Foundations of the Theory of Probability* (Segunda Edición ed.). New York: Chelsea Publishing Company.

Kotz, S., & van Dorp, J. (2004). *Beyond Beta: Other Continuous Families Of Distributions With Bounded Support And Applications.* Singapore: World Scientific Publishing Co. Pte. Ltd.

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications.* Boca Raton: Taylor & Francis Group, LLC.

Laplace, P.-S. (2015). *Ensayo Filosófico Sobre Probabilidades.* Ciudad de México: Biblioteca Digital del Instituto Latinoamericano de Comunicación Educativa. Retrieved from http://bibliotecadigital.ilce.edu.mx/Colecciones/ReinaCiencias/_docs/EnsayoFilosoficoProbabilidades.pdf

Loughborough University. (2008, Febrero 21). *Total Probability and Bayes' Theorem.* Retrieved from The theorem of total probability: https://learn.lboro.ac.uk/archive/olmp/olmp_resources/pages/workbooks_1_50_jan2008/Workbook35/35_4_total_prob_bayes_thm.pdf

Maibaum, G. (1988). *Teoría de Probabilidades y Estadística Matemática.* (M. Á. Pérez, Trans.) La Habana, Cuba: Editorial Pueblo y Educación.

Marx, K. (2007). *Elementos Fundamentales para la Crítica de la Economía Política (Grundrisse) 1857-1858* (Vol. I). México, D.F.: Siglo XXI.

MIT. (2017, Agosto 14). *R Documentation*. Retrieved from The Negative Binomial Distribution: https://web.mit.edu/~r/current/arch/i386_linux26/lib/R/library/stats/html/NegBinomial.html#:~:text=A%20negative%20binomial%20distribution%20can,%2Dinteger%20values%20of%20size%20.)

Mittelhammer, R. (2013). *Mathematical Statistics for Economics and Business* (Segunda ed.). New York: Springer.

North Carolina State University. (2020, Septiembre 27). *People - Department of History* . Retrieved from Dr Edith D Sylla: https://history.ncsu.edu/people/faculty_staff/edsssl

Perezgonzalez, J. (2015, Marzo 3). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. (L. Roberts, Ed.) *Frontiers in Psychology, 6*(223), 1-11. doi:10.3389/fpsyg.2015.00223

Poisson, S.-D. (2013). *Researches into the Probabilities of Judgments in Criminal and Civil Cases.* (O. Sheynin, Ed.) Berlin: arXiv. Retrieved from https://arxiv.org/abs/1902.02782

Radboud Univeristy. (2011, Febrero 11). *Faculty of Philosophy, Theology and Religious Studies*. Retrieved from Center for the History of Philosophy and Science. Edith Dudley Sylla: https://www.ru.nl/ptrs/chps/about-us/former-members/vm/sylla/

Rosental, M. M., & Iudin, P. F. (1971). *DICCIONARIO FILOSÓFICO.* San Salvador: Ediciones Tecolut.

Russell, K. (2014, Enero 29). *University of Manitoba*. Retrieved from Hypothesis testing: http://home.cc.umanitoba.ca/~krussll/stats/hypothesis-testing.html

Smithson, M., & Verkuilen, J. (2006). *Fuzzy Set Theory (Applications in the Social Sciences).* California: Sage Publications Inc.

Stack Exchange. (2015, Mayo 14). *Mathematics*. Retrieved from Why is a geometric progression called so?: https://math.stackexchange.com/questions/1281856/why-is-a-geometric-progression-called-so

Stack Exchange. (2020, Septiembre 30). *Cross Validated*. Retrieved from Negative binomial distribution vs binomial distribution: https://stats.stackexchange.com/questions/176034/negative-binomial-distribution-vs-binomial-distribution

Student. (1907). On the Error of Counting with a Haemacytometer. *Biometrika*, 351-360.

Swat, M., Grenon, P., & Wimalaratne, S. (2017). *ProbOnto 2.5: Ontology and Knowledge Base of Probability Distributions.* ProbOnto.

Taboga, M. (2012). *Lectures On Probability Theory and Mathematical Statistics.* California: CreateSpace Independent Publishing Platform.

Tamir, D. E., Rishe, N. D., & Kandel, A. (2015). *Fifty Years of Fuzzy Logic and its Applications.* New York: Springer.

Vucinich, A. (1982). Soviet Marxism and the History of Science. *The Russian Review, 41*(2), 123-143.

Wackerly, D., Mendelhall III, W., & Scheaffer, R. (2010). *Estadística Matemática con Aplicaciones* (Séptima ed.). (J. H. Muñoz, Trans.) México. D.F.: Cengage Learning Editores, S.A.

Walck, C. (1996). *Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists (Internal Report SUF–PFY/96–01)*. Stockholm: Particle Physics Group of University of Stockholm.

Weisstein, E. (2020, Septiembre 22). *MathWorld*. Retrieved from Negative Binomial Series: https://mathworld.wolfram.com/NegativeBinomialSeries.html

Weisstein, E. (2020, Septiembre 29). *Wolfram MathWorld*. Retrieved from Negative Binomial Distribution: https://mathworld.wolfram.com/NegativeBinomialDistribution.html

Weisstein, E. (2020, Septiembre 30). *Wolfram MathWorld*. Retrieved Septiembre 30, 2020, from Pascal Distribution: https://mathworld.wolfram.com/PascalDistribution.html

Wikipedia. (2020, Septiembre 22). *Negative Binomial Distribution*. Retrieved from Probability mass function: https://en.wikipedia.org/wiki/Negative_binomial_distribution

Wikipedia. (2020, Septiembre 27). *Population Genetics*. Retrieved from J. B. S. Haldane: https://es.wikipedia.org/wiki/John_Burdon_Sanderson_Haldane

Wikipedia. (2020, Septiembre 20). *Statistics*. Retrieved from Kernel: https://en.wikipedia.org/wiki/Kernel_%28statistics%29

Wikipedia. (2020, Septiembre 24). *Statistics*. Retrieved from Degrees of Freedom: https://en.wikipedia.org/wiki/Degrees_of_freedom_(statistics)

Wikipedia. (2020, Septiembre 22). *Statistics*. Retrieved from Negative Binomial Distribution: https://en.wikipedia.org/wiki/Negative_binomial_distribution

Wikipedia. (2020, Septiembre 23). *Statistics*. Retrieved from Inverse Probability: https://en.wikipedia.org/wiki/Inverse_probability

Wikipedia. (2020, Septiembre 13). *statistiques*. Retrieved from Noyau : https://fr.wikipedia.org/wiki/Noyau_(statistiques)

Wikipedia. (2020, Septiembre 15). *Topology*. Retrieved from Manifold Decomposition: https://en.wikipedia.org/wiki/Manifold_decomposition

Williamson, J. (2010). *In Defence of Objective Bayesianism.* Oxford: Oxford University Press.

Zadeh, L. A. (1975). *Fuzzy Sets and Their Applications to Cognitive and Decision Processes.* (K.-S. Fu, K. Tanaka, & M. Shimura, Eds.) California: Academic Press Inc.

Википедия. (2020, Septiembre 13). *статистика*. Retrieved from Ядро: https://ru.wikipedia.org/wiki/%D0%AF%D0%B4%D1%80%D0%BE_(%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0)