# 7
# Pruebas repetidas

En este capítulo se tratan dos tipos de procesos básicos en pruebas repetidas. El primero es el estudiado por Bernoulli en Ginebra, a fines del siglo XVI, aplicado a los juegos de azar. Condujo al modelo de la Probabilidad Binomial, Binomial Negativa, Probabilidad de Pascal y la Probabilidad Multinomial. El segundo fue desarrollado por Poisson en París, en los inicios del siglo XVIII, como un caso particular de la binomial aplicado a los casos raros. La Probabilidad de Poisson se aplica en las técnicas de recuentos celulares, fenómenos de contagio, virología, como así también en la Ley de Radiactividad General, seguros de vida, teoría de las líneas de espera o teoría de colas, etc.

## 7.1 Procesos tipo Bernoulli

Se denominan procesos de tipo Bernoulli, a todo experimento consistente en una serie de pruebas repetidas, caracterizadas por tener resultados que se pueden clasificar en si verifican o no cierta propiedad o atributo, siendo aleatorios e independientes. La relación con un proceso de tipo hipergeométrico, visto en el capítulo anterior, es directa: Si las extracciones se hacen *con reposición* se transforma en un proceso del tipo Bernoulli.

Para identificar un proceso Bernoulli en una serie de pruebas repetidas, se deben verificar tres condiciones:

1) *Resultados dicotómicos*: Los resultados de cada prueba se pueden clasificar en "éxito" si verifican cierta condición, o "fracaso" en el caso contrario. No importa el tipo de magnitud clínica que se trate; al efectuarle la determinación en el laboratorio, siempre se podrá ver si el resultado obtenido en la muestra del paciente verifica o no cierto atributo. Y entonces, siempre se podrán clasificar los resultados en forma dicotómica. En el caso de la determinación de la CPK -que es una magnitud de tipo continua- al tomar un punto de corte se clasifican los infinitos resultados posibles en dos: Infartado - No infartado. En una determinación del fenómeno Psi, el recuento de resultados -que es una magnitud de tipo discreta- siempre se podrá clasificar en: hay evidencia de percepción extrasensorial, o no la hay. En un concurso de belleza, o el resultado de una carrera (magnitudes ordinales), se puede subdividir en: ganó-perdió.

2) *Independencia de las pruebas*: El resultado de una prueba cualquiera es independiente del resultado obtenido en la prueba anterior, y no incide en el resultado de la prueba siguiente. Esto se cumple en los juegos de azar como al lanzar una moneda o un dado, y en la extracción de barajas, si se realizan con reposición luego de cada extracción. Es uno de los supuestos básicos de la Teoría Genética de Mendel y en las leyes de la herencia humana, como el sexo de un recién nacido. La independencia deja de cumplirse cuando se da un fenómeno de contagio o repulsión.

3) *Estabilidad de las pruebas*: La probabilidad p de obtener un resultado considerado como un éxito se mantiene constante a lo largo de toda la serie de pruebas. Análogamente, para la probabilidad q de obtener un fracaso. Como los resultados son dicotómicos, el universo de resultados aparece particionado por ambos sucesos, entonces $p + q = 1 = P(\boldsymbol{S})$.

La tirada de una moneda o de un dado es el ejemplo clásico de este proceso. Se debe notar que eso no cambia si la moneda o el dado están viciados. Si un dado cargado tiene una probabilidad de 2/6 de sacar un as, como esa probabilidad se mantiene constante e independiente a lo largo de la serie de pruebas, se cumplen las condiciones anteriores.

En cambio, si se da un fenómeno de reacción en cadena, o en cascada, o contagio, las probabilidades dejan de ser constantes. Por ejemplo, si la probabilidad de enfermarse de hepatitis en Misiones es p, y se presenta una epidemia, entonces el contagio hace crecer mucho el número de enfermos, p cambia y deja de ser estable. El corrimiento de los puntos de una media de nylon femenina es un ejemplo de fenómeno en cascada. Para los últimos casos, se podrá emplear otro modelo, el de Poisson, que se verá más adelante. Notar que el Odds de éxitos en el modelo Bernoulli es $p / q = p / (1-p)$.

# 7.2 Probabilidad Binomial

Cuando en un proceso del tipo Bernoulli se desea saber la probabilidad de obtener exactamente r éxitos, en una serie de n pruebas, con una probabilidad de éxito p, se puede aplicar la fórmula de la probabilidad binomial:

$$\boxed{P_B\ (r / n , p) = C(n , r) \,.\, p^r\ (1 - p)^{n-r}}$$

Sea $S_1$ la primera serie de n pruebas realizadas donde se verifica la ocurrencia de r éxitos, con una probabilidad p y la de fracaso q:

$\boldsymbol{S_1}$ **:** $\boldsymbol{E\ F\ E\ F\ F\ ....E}$ (donde $\boldsymbol{E}$ indica la ocurrencia de un éxito y $\boldsymbol{F}$ la de un fracaso)

$P(\boldsymbol{S_1}) = P(\boldsymbol{E} \cap \boldsymbol{F} \cap \boldsymbol{E} \cap \boldsymbol{F} \cap \boldsymbol{F} \cap ... \cap \boldsymbol{E})$ y como los sucesos son independientes

$P(\boldsymbol{S_1}) = P(\boldsymbol{E}) \,.\, P(\boldsymbol{F}) \,.\, P(\boldsymbol{E}) \,.\, P(\boldsymbol{F}) \,.\, P(\boldsymbol{F})\ ...\ P(\boldsymbol{E}) = p\,.q\,.p\,.\,q\,.\,q\ ...\ p = p^r\ (1 - p)^{n-r}$

Pero esta secuencia, es una de todas las k posibles series donde se verifican r éxitos en n pruebas. Otra serie cualquiera como la $\boldsymbol{S_2}$ **:** $\boldsymbol{E\ F\ E\ F\ F\ ....F}$

$P(\boldsymbol{S_2}) = P(\boldsymbol{E} \cap \boldsymbol{F} \cap \boldsymbol{E} \cap \boldsymbol{F} \cap \boldsymbol{F} \cap ... \cap \boldsymbol{F}) = P(\boldsymbol{E}) \,.\, P(\boldsymbol{F}) \,.\, P(\boldsymbol{E}) \,.\, P(\boldsymbol{F}) \,.\, P(\boldsymbol{F})\ ...\ P(\boldsymbol{F})$
$P(\boldsymbol{S_2}) = p\,.q\,.p\,.\,q\,.\,q\ ...\ q = p^r\ (1 - p)^{n-r}$

En esta segunda serie, diferente de la primera en el orden de ocurrencia de los sucesos, se obtiene la misma probabilidad. Generalizando, para una serie cualquiera i es: $P(\boldsymbol{S_i}) = p^r\ (1 - p)^{n-r}$

El número k total de pruebas posibles se obtiene con la combinatoria de n elementos tomados de a r. O sea: C (n , r). La probabilidad binomial pedida será la probabilidad de la unión de todos los sucesos posibles, esto es :

$$P_B\ (r / n, p) = P(\boldsymbol{S_1} \cup \boldsymbol{S_2} \cup \boldsymbol{S_3} \cup \ldots \cup \boldsymbol{S_k}) = P(\cup_k\ \boldsymbol{S_k}) = \sum_k P(\boldsymbol{S_k}) = k \cdot P(\boldsymbol{S_k})$$

Y reemplazando por las relaciones anteriores, se llega a la fórmula de la Binomial de más arriba

*Ejemplo 1*) Sea el caso de una droga X, con una dosis mortal de 1g/100 ml para cobayos experimentales, en el 25% de los casos. Aplicando esta dosis a cien cobayos se desea saber cuanto vale la probabilidad de que mueran veinte de ellos.

Esto se calcula con la binomial, siempre y cuando se cumplan los supuestos básicos, a saber:

1) Los cobayos mueren (éxito) o sobreviven (fracaso).
2) Que un cobayo muera con la dosis, no significa que lo hará el siguiente (independencia) pues no se trata de una epidemia.
3) La probabilidad de que mueran se mantiene constante a lo largo de la serie de pruebas (p = 0,25). Entonces, aplicando la fórmula será:

$$P_B\ (r=20 / n=100, p=0{,}25) = C(100, 20) \cdot (0{,}25)^{20}\ (1 - 0{,}25)^{80} = 0{,}04902$$

*Ejemplo 2*) En el Cuadro 7.1 se presenta un ejemplo de proporción de sexos la descendencia de parejas humanas; los datos provienen de un estudio de Geissler (1889), quien analizó 6.115 familias de doce hijos cada una en Sajonia. La proporción de sexos es 1 : 1 aproximadamente. Sin embargo, como se sabe que en poblaciones humanas diferentes estos pueden variar un poco, para este trabajo se tomó la proporción obtenida en el mismo. La proporción de 519 : 481 entre hombres y mujeres. Si el número de hijos es n = 12, la cantidad r de mujeres, puede ser imaginada como una variable que va de 0 a 12, ya para cada valor de r habrá un valor de la binomial que se muestra en la columna $P_B$ ( r ) del Cuadro 7.1, calculado con :

$$P_B\ (r=0 / n=12, p=0{,}481) = C(12, 0) \cdot (0{,}481)^{0}\ (0{,}519)^{12} = 0{,}00038$$

Eso implica que en 6.115 familias de 12 hijos, la probabilidad de encontrar familias sin hijas mujeres es de 0,00038, esto es, una frecuencia esperada de 2,3 familias del total. Los valores esperados se obtienen multiplicando el número total de familias estudiadas por la frecuencia esperada. Esto es, 6.115 $P_B$ ( r = 12) = 2,3. Geissler obtuvo 7 casos reales, más del doble de lo esperado. En el caso inverso, es decir familias de 12 hijas, da : 6.115 $P_B$ ( r = 0) = 0,9 familias es el número de casos esperado. Pero hubo 3 casos reales, casi el triple. Si se comparan los valores esperados con los observados, se nota una buena concordancia en los valores del centro de la tabla, pero en las puntas eso deja de verse. Por ejemplo, para el caso de 6 mujeres y 6 varones, se observan 1.343 casos y se esperaban 1.367,3 o sea una diferencia del orden del 1,8%; en cambio, en ambas puntas ronda el 100%. La base genética de este fenómeno no está muy clara. Pero hay familias que "tienden" a tener todos varones y otras a mujeres. Todo ocurre como si hubiese una especie de "contagio". Para tener una prueba científica de esto, hay que realizar un *test estadístico* para hacer una prueba de hipótesis, como se verá en capítulos posteriores. En muchos casos de estudios mendelianos, la proporción de sexos se calcula con la relación 1:1; en este caso no se hizo así, sino que se usaron los *datos experimentales* para su obtención, en lugar de los *teóricos*. Más adelante, se verá la importancia de esta diferencia.

**Cuadro 7.1 Aplicaciones de la Probabilidad Binomial.**

*Caso 1 :* Proporción de sexos en 6.115 familias con doce hijos de Sajonia -datos tomados de Sokal(*)

| Cantidad de mujeres en familias de 12 hijos r | Probabilidad Binomial $P_B$ ( r ) | Número de casos esperados $Er = N . P_B$ ( r ) | Número de casos observados Or | Diferencias relativas (Or - Er)/Er |
|---|---|---|---|---|
| 0 | 0,000384 | 2,3 | 7 | + 2,04 |
| 1 | 0,004264 | 26,1 | 45 | + 0,724 |
| 2 | 0,021725 | 132,8 | 181 | + 0,363 |
| 3 | 0,067041 | 410,0 | 478 | + 0,166 |
| 4 | 0,139703 | 854,3 | 829 | - 0,030 |
| 5 | 0,206973 | 1265,6 | 1112 | - 0,121 |
| 6 | 0,223590 | 1367,3 | 1343 | - 0,018 |
| 7 | 0,177459 | 1085,2 | 1033 | - 0,048 |
| 8 | 0,102708 | 628,1 | 670 | + 0,066 |
| 9 | 0,042280 | 258,5 | 286 | + 0,106 |
| 10 | 0,011743 | 71,8 | 104 | + 0,448 |
| 11 | 0,001975 | 12,1 | 24 | + 0,164 |
| 12 | 0,000153 | 0,9 | 3 | + 2,333 |

*Ejemplo 3*) Se toman al azar 1000 muestras de n = 5 individuos cada una, de la especie Homo Sapiens. Se formula la hipótesis de un 40% de ellos presenta cierta enfermedad, no contagiosa. Con computadora se simulan los casos observados usando la tabla de números al azar - (*)

| Cantidad de enfermos por muestra r | Probabilidad Binomial $P_B$ ( r ) | Número de casos esperados $Er = N . P_B$ ( r ) | Número de casos simulados Or | Diferencias relativas (Or - Er)/Er |
|---|---|---|---|---|
| 0 | 0,07776 | 388,8 | 416 | + 0,0694 |
| 1 | 0,25920 | 1296,0 | 1327 | + 0,0238 |
| 2 | 0,34560 | 1728,0 | 1686 | - 0,0243 |
| 3 | 0,23040 | 1152,0 | 1110 | - 0,0365 |
| 4 | 0,07680 | 384,0 | 406 | + 0,0586 |
| 5 | 0,01024 | 51,2 | 55 | + 0,0703 |

En este ejemplo, se simula la población humana, con la tabla de números al azar, haciendo N= 1000 extracciones de 5 números cada una; como no hay límite para hacer esto, el tamaño teórico de la población es infinito. Cuando en la muestra aparecen los números 1,2,3 o 4 se toma como infectada. Cuando aparecen los números 0,5,6,7,8 o 9 se la toma como no infectada. Por ejemplo, si en una de las muestras simuladas resulta (3,8,9,6,0) se la cuenta como una muestra con 1 infectado. Estos datos se vuelcan en la columna correspondiente a frecuencias observadas. En la primer columna se calcula la probabilidad binomial con un número de éxitos que varía de 0 a 5, un tamaño de muestra de n = 5, y una probabilidad p = 0,4. Esta porque hay diez dígitos posibles en la tabla de números al azar, de los cuales 4 se usan para los éxitos. Como puede verse hay una buena concordancia (ninguna diferencia relativa supera el 7% ) entre los valores esperados y los observados. Se cumple que cada muestra es independiente de la otra y que la probabilidad es constante a lo largo de la serie de pruebas.

*Ejemplo 4*) En una farmacia se ha calculado la probabilidad de venderle a un cliente con obra social es del 20%. Se eligen al azar 15 clientes de ese tipo que ingresan al negocio y se desea calcular la probabilidad de concretar menos de tres ventas.

Para analizar este problema conviene determinar todos los casos posibles. Menos de 3 ventas significa que se pueden concretar dos, una o ninguna venta, es decir los casos posibles son:

***A*** : no se vendió nada ( r = 0 )
***B*** : solo se hizo una venta ( r = 1)
***C*** : se concretaron dos ventas ( r = 2)

Estos sucesos son excluyentes entre sí, por lo tanto aplicando el Axioma 3 generalizado será:

$$P(\text{menos de 3 ventas}) = P(\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{C}) = P(\boldsymbol{A}) + P(\boldsymbol{B}) + P(\boldsymbol{C})$$

$$P(\boldsymbol{A}) = P_B\ (r=0 / n=15, p=0{,}2) = C(15, 0)\ .\ (0{,}2)^0\ (0{,}8)^{15} = 0{,}0352$$
$$P(\boldsymbol{B}) = P_B\ (r=1 / n=15, p=0{,}2) = C(15, 1)\ .\ (0{,}2)^1\ (0{,}8)^{14} = 0{,}1319$$
$$P(\boldsymbol{C}) = P_B\ (r=2 / n=15, p=0{,}2) = C(15, 2)\ .\ (0{,}2)^2\ (0{,}8)^{13} = 0{,}2309$$

Sumando los tres valores anteriores se obtiene P (menos de 3 ventas) = 0,398. Es decir, hay casi un 61% de probabilidades de que se concreten 3 ventas o más.

# 7.2.1 Contagio y repulsión

Para aplicar la fórmula de la función binomial se deben verificar los tres supuestos de un proceso Bernoulli. El ejemplo clásico es en los juegos de azar. Recordando el problema de lanzar tres veces una moneda, resuelto con el diagrama del árbol en el punto 6.1, puede verse que es mucho más elegante y simple resolverlo con la binomial. En efecto, la probabilidad de sacar tres caras (o secas) seguidas será:

$$P_B\ (r=3 / n=3, p=0{,}5) = C(3, 3)\ .\ (0{,}5)^3\ (0{,}5)^0 = (0{,}5)^3 = 0{,}125$$

Naturalmente, si en lugar de ser tres, fueran cien lanzamientos, la resolución usando el Diagrama del árbol sería sumamente engorrosa y es más práctico usar este modelo.

En los lanzamientos de una moneda, es muy claro el cumplimiento de los tres supuestos, pero en los casos de la vida real, ya no lo son tanto. Si se piensa en el ejemplo visto cuando se simuló por computadora el número de casos observados usando una tabla de números al azar, los valores necesariamente son aleatorios e independientes. Pero cuando dejan de cumplirse uno o más de estos supuestos, los resultados varían mucho. Sea el mismo caso anterior, pero ahora se le obliga a la computadora a seguir buscando en la tabla, cada vez que salga el valor 3, hasta que encuentre otro de los cuatro números correspondientes al suceso: infectado. Entonces, las probabilidades de cada número dejan de ser iguales, no son más constantes. Luego de efectuar un cierto número de sorteos aumentará la frecuencia de las clases con dos o más enfermos en la muestra y disminuirá el número de clases con un solo enfermo. Todo ocurre como si, cuanto más enfermos hay en la muestra, más fácil se enferman sus vecinos. O sea, se ha simulado un fenómeno de *contagio.* En cambio, si se arregla el programa de la computadora, para que sean más frecuentes los casos heterogéneos, es decir, con similar proporción de infectados, disminuyen los casos más homogéneos (todos sanos, o todos infectados) y se simula un fenómeno de *repulsión*.

**Cuadro 7.2 Contagio y repulsión**

| Cantidad de enfermos por muestra r | Probabilidad Binomial $P_B\ (r)$ | Número de casos esperados $Er = N \cdot P_B\ (r)$ | Número de casos simulados para *contagio* | Número de casos simulados para *repulsión* |
|---|---|---|---|---|
| 0 | 0,07776 | 388,8 | 464 | 295 |
| 1 | 0,25920 | 1296,0 | 1450 | 1275 |
| 2 | 0,34560 | 1728,0 | 1368 | 1946 |
| 3 | 0,23040 | 1152,0 | 1160 | 1131 |
| 4 | 0,07680 | 384,0 | 488 | 324 |
| 5 | 0,01024 | 51,2 | 70 | 29 |

En el Cuadro 7.2 se muestran los resultados de la simulación. Las frecuencias de *contagio* aumentan mucho para los valores extremos de r y disminuyen para r=2. Lo contrario ocurre en la repulsión, con dos enfermos por muestra la frecuencia sube y baja en los dos extremos. La forma correcta de verificar si estas diferencias son debidas al efecto de contagio o repulsión, es con una prueba o test de hipótesis estadístico, que se verá más adelante.

Si en vez de ser un caso simulado, fuese real, el significado de un *contagio* es que cada vez se encuentra más muestras totalmente infectadas o totalmente sanas, y menos muestras con igualdad de sanos y enfermos. Todo esto, con respecto a lo que cabría esperar si la probabilidad de contraer la enfermedad fuese *independiente*. El contagio es un fenómeno entre los integrantes de cada muestra. O sea, si uno de ellos está enfermo, la probabilidad de enfermarse aumenta en sus "vecinos" de la muestra. Cabe la posibilidad de encontrarse con este fenómeno en la práctica debido a una mala elección de las muestras, porque el investigador tiene tendencia a tomar muestras infectadas. Es decir, no hay un contagio sino una selección mal hecha. Pero si se hace correctamente, con un muestreo al azar, entonces los resultados muestran contagio. Por ejemplo, puede haber ocurrido que se hayan expuesto a un mismo foco infeccioso, o que se hayan contagiado en el intervalo de tiempo entre que fueron extraídos y revisados.

El fenómeno opuesto, ya es más complicado para explicarlo clínicamente. El número de grupos homogéneos ha disminuido sensiblemente, aumentando aquellos grupos con proporciones similares de enfermos y no enfermos. Todo ocurre como si, al estar dos o tres individuos de la muestra enfermos, los restantes son menos propensos a estarlo. Como si los enfermos de alguna manera inmunizaran a sus compañeros muestrales. Se da una especie de *repulsión*. Una interpretación es que hay un número limitado de gérmenes patógenos y una vez que se ubicaron en sus respectivos huéspedes, se agotó la fuente de infección. En infecciones del tipo microbianas, esto es menos plausible, pero muy razonable en el caso de parásitos. Otra cosa que puede haber ocurrido es que con los primeros enfermos se haya vacunado a sus vecinos y de allí los resultados.

# 7.3 Probabilidad Pascal

Cuando en un proceso Bernoulli, se busca la probabilidad de que sean necesarias exactamente n pruebas, para lograr r éxitos, con una probabilidad de éxito p, entonces se recurre a la fórmula de Pascal, para resolver el problema.

$$\boxed{P_{Pa}\ ( n / r , p ) \ = \ C\,(n-1 , r-1)\ .\ p^r\ (1 - p)^{n-r}}$$

Sea $S_1$ la primera serie de n pruebas realizadas donde se verifica la ocurrencia del r-avo éxito, en la prueba número n, con una probabilidad p y la de fracaso q:

$\boldsymbol{S_1}$ **:** ***E F E F F ....F E*** (donde el último debe ser un hay ***E*** )

$P(\boldsymbol{S_1}) = P(\boldsymbol{E} \cap \boldsymbol{F} \cap \boldsymbol{E} \cap \boldsymbol{F} \cap \boldsymbol{F} \cap ... \cap \boldsymbol{F} \cap \boldsymbol{E}) = P(\boldsymbol{R_1} \cap \boldsymbol{E}) = P(\boldsymbol{R_1}) . P(\boldsymbol{E})$

Donde $\boldsymbol{R_1}$ es el suceso : Obtención de r-1 éxitos en las primeras n-1 pruebas realizadas dentro de la primer serie de pruebas $\boldsymbol{S_1}$. Como los sucesos son independientes se obtiene

$P(\boldsymbol{S_1}) = P(\boldsymbol{E}) . P(\boldsymbol{F}) . P(\boldsymbol{E}) . P(\boldsymbol{F}) . P(\boldsymbol{F}) ... P(\boldsymbol{F}) . P(\boldsymbol{E}) = p .q .p . q . q ... p = p^r\ (1 - p)^{n-r}$

Pero esta secuencia es una de todas las m posibles series donde se verifican r éxitos en n pruebas. Otra serie cualquiera como la $\boldsymbol{S_2}$ **:** ***F F E F F ....E***

$P(\boldsymbol{S_2}) = P(\boldsymbol{F} \cap \boldsymbol{F} \cap \boldsymbol{E} \cap \boldsymbol{F} \cap \boldsymbol{F} \cap ... \cap \boldsymbol{E}) = P(\boldsymbol{F}) . P(\boldsymbol{F}) . P(\boldsymbol{E}) . P(\boldsymbol{F}) . P(\boldsymbol{F}) ... P(\boldsymbol{E})$
$P(\boldsymbol{S_2}) = P(\boldsymbol{R_2} \cap \boldsymbol{E}) = P(\boldsymbol{R_2}) . P(\boldsymbol{E}) = p .q .p . q . q ... q = p^r\ (1 - p)^{n-r}$

En esta segunda serie, diferente de la primera en el orden de ocurrencia de los sucesos, se obtiene la misma probabilidad. Generalizando, para una serie i es :

$$P(\boldsymbol{S_i}) = p^r\ (1 - p)^{n-r}$$

El número $m$ total de pruebas posibles se obtiene, con la combinatoria de n-1 elementos tomados de a r-1 : C (n-1 , r-1). La probabilidad Pascal pedida, será la probabilidad de la unión de todos los sucesos posibles, esto es:

$$P_{Pa}\ ( r / n , p ) = P(\boldsymbol{S_1} \cup \boldsymbol{S_2} \cup \boldsymbol{S_3} \cup ... \cup \boldsymbol{S_m}) = P(\cup_{\boldsymbol{m}}\ \boldsymbol{S_m}) = \sum_{\boldsymbol{m}} P(\boldsymbol{S_m}) = m . P(\boldsymbol{S_m})$$

Reemplazando m con el valor de la combinatoria, se llega a la fórmula general de la $P_{Pa.}$

Otra forma de deducir lo mismo es la siguiente:

Sea ***R*** el suceso: Obtención de r-1 éxitos en las primeras n-1 pruebas realizadas
y ***X*** el suceso: Obtención de un éxito en la última prueba realizada

**J.C. Azzimonti Renzo: arroi_pss@ciudad.com.ar**

Entonces la probabilidad buscada será

$P_{Pa}$ ( n / r , p ) = P (***R*** ∩ ***X*** ) = P (***R***) . P (***X***)

Donde P (***R***) es una binomial de parámetros r-1, n-1 y P (***X***) es la probabilidad p de éxito

$P_{Pa}\ (n / r , p) = P_B\ (r-1 / n-1 , p) . p = C\ (n-1 , r-1) .\ p^{r-1}\ (1 - p)^{n-1-r+1} . p$

$$P_{Pa}\ (n / r , p) = C\ (n-1 , r-1) .\ p^{r}\ (1 - p)^{n-r}$$

La relación entre la binomial y la Pascal se deduce con:

$$C(n-1 , r-1) = (r/n) .\ C(n , r)$$

De donde,

$$\boxed{P_{Pa}\ (n / r , p) = (r/n) . P_B\ (r / n , p)}$$

Lo que significa que todos los problemas del tipo Pascal se pueden resolver con la Binomial. Desde el punto de vista del cálculo si se tienen tablas para los valores binomiales (ver la bibliografía), no hace falta tener otra tabla para las probabilidades tipo Pascal.

*Ejemplo 1*) Se debe calcular la probabilidad de que sean necesarios 100 ensayos para lograr la vigésima muerte de los cobayos experimentales, del primer ejemplo del tema 7.2.

Aplicando la fórmula general es:

$P_{Pa}\ (n=100 / r=20 , p=0{,}25) = C\ (99 , 19) .\ (0{,}25)^{20}\ (0{,}75)^{80} = 0{,}00986$

Usando los valores calculados antes con la binomial resulta lo mismo:

$P_{Pa}\ (100 / 20 , 0{,}25) = (20/100)\ P_B\ (r=20/n=100, p=0{,}25) = 0{,}2 . 0{,}0493 = 0{,}00986$

*Ejemplo 2*) Usando los datos del cuarto ejemplo del punto 7.2 calcular la probabilidad de que sean necesarios 15 intentos antes de concretar la primera venta.

$P_{Pa}\ (n=15 / r=1 , p=0{,}2) = (1/15)\ P_B\ (r=1 / n=15 , p=0{,}2) = (0{,}067) .\ C\ (15 , 1) .\ (0{,}2)^{1} . (0{,}8)^{14}$

$P_{Pa}\ (n=15 / r=1 , p=0{,}2) = (0{,}067)\ (0{,}1319) = 0{,}0088$

*Ejemplo 3*) La probabilidad de que un alumno elegido al azar apruebe un examen es de 0,6. Si se eligen 5 alumnos al azar calcular la probabilidad de que haya que corregir 3 exámenes antes de tener al primer aprobado.

$P_{Pa}\ (n=5 / r=3 , p=0{,}6) = (3/5)\ P_B\ (r=3 / n=5 , p=0{,}6) = (0{,}6). C(5,3)\ (0{,}6)^{3}(0{,}4)^{2} = 0{,}21$

# 7.4 Probabilidad Binomial Negativa

Cuando en un proceso de tipo Bernoulli se desea saber la probabilidad de que sean necesarias efectuar exactamente g pruebas adicionales a fin de obtener r éxitos, en un total de n pruebas, se puede aplicar la fórmula de la binomial negativa para calcularla:

$$P_{BN}\ ( g / n, r , p ) = C\ (g+r-1\ , r-1)\ .\ p^r\ (1 - p)^g$$

Donde g = n-r pues las pruebas adicionales a r éxitos solo pueden ser fracasos. Entonces, el número de pruebas será n = g + r . La probabilidad pedida es del tipo Pascal , pero para g fracasos; por lo tanto con un simple reemplazo se puede llegar a lo buscado:

$$P_{BN}\ (g / n, r , q)=P_{Pa}( n=g+r / r , q)=C\ (g+r-1\ , r-1).(1-q)^r\ (q)^{g+r-r} = C\ (g+r-1\ , r-1)\ .\ p^r\ (1 - p)^g$$

Otra manera de expresarla más tradicional es la siguiente. Sea q = 1 - p , la probabilidad de un fracaso al hacer el experimento. Sean además Q = (1/p) y P = (q/p), entonces se pueden deducir las relaciones siguientes: (Q-P) = 1 y (1-p) = (P/Q) reemplazando en la fórmula de Pascal será:

$$P_{Pa}\ ( n / r , q ) = C\ (g+r-1\ , r-1)\ .\ Q^{-r}\ (P/Q)^g = P_{BN}\ ( g / n, r , q )$$

Siendo esta última la manera convencional de expresar la probabilidad binomial negativa. Sin embargo, se debe destacar que se trata de otra manera de ver la probabilidad de Pascal.

*Ejemplo 1*) Sea un sanatorio al cual arriban pacientes a tratarse diversas afecciones. Un 40% de los mismos concurren a efectuarse análisis clínicos. ¿ Cuánto vale la probabilidad de que veinte pacientes sigan de largo, antes de que el quinto paciente entre al laboratorio a tratarse, si se supone independencia en la llegada?

De acuerdo a los datos es : p = 0,4 ; q = 0,6 ; r =5 y g = 20 luego el valor pedido será :

$$P_{BN}\ ( g / n, r , p ) = C\ (20+5-1\ , 5-1)\ .\ 0{,}4^5\ (0{,}6)^{20} = 0{,}004$$

*Ejemplo 2*) Sea una farmacia donde el 40% de los clientes compran artículos de perfumería. Calcular la probabilidad de que 20 clientes no se acerquen a dicho sector, antes de que el cliente número 25 pida un artículo de perfumería.

Los datos son los mismos del problema anterior, por lo que la respuesta es:

$$P_{BN}\ ( 20 / 25\ , 5\ , 0{,}4 ) = 0{,}004$$

# 7.5 Probabilidad Geométrica

Cuando en un proceso de tipo Bernoulli se desea saber la probabilidad de que sean necesarias efectuar exactamente g pruebas antes de obtener el primer éxito (r = 1), en un total de n pruebas, donde = g +1. Se puede aplicar la fórmula de la probabilidad geométrica:

$$\boxed{P_G\ (g/p) = p\cdot q^{g}}$$

*Ejemplo 1*) Si en una farmacia la probabilidad de que un cliente compre un artículo de perfumería es del 30%, un remedio del 60%, y cualquier otra cosa un 10%. Calcular la probabilidad de tener que hacer 4 intentos antes de vender el primer: a) remedio; b) un artículo de perfumería.

a) Para remedio es : $P_G\ (g/p) = p\cdot q^{g} = (0{,}6)\cdot(0{,}4)^4 = 0{,}01536$

b) Para perfumería es: $P_G\ (g/p) = (0{,}3)\cdot(0{,}7)^4 = 0{,}07203$

*Ejemplo 2*) Si en un laboratorio de análisis clínicos especializado en HIV la probabilidad de que un paciente de positivo es del 20%. Calcular la probabilidad de que entren 8 pacientes antes de encontrar al primer positivo.

$P_G\ (g/p) = p\cdot q^{g} = (0{,}2)\cdot(0{,}8)^8 = 0{,}034$

# 7.6 Probabilidad Multinomial

Sean los sucesos ***C₁, C₂, C₃, ..., Cₖ*** todos mutuamente excluyentes entre sí, cuyas respectivas probabilidades de ocurrencia son $P(C_1)$, $P(C_2)$ .... $P(C_k)$. Entonces, la probabilidad de que en n pruebas sucesivas el suceso $C_1$ ocurra $r_1$ veces, el $C_2$ suceda $r_2$ veces, .... y el suceso $C_k$ ocurra $r_k$ veces, se puede calcular con la fórmula de la *probabilidad multinomial*:

$$\boxed{P_M = \frac{n!}{r_1!\cdot r_2!\ \ldots\ r_k!}\cdot P(C_1)^{r_1}\cdot P(C_2)^{r_2}\ \ldots\ P(C_k)^{r_k}}$$

Esta probabilidad es una generalización de la probabilidad binomial, donde $r_1 + r_2 + \ldots + r_n = n$
Y en el caso particular que k=2, la multinomial se reduce a la binomial.

*Ejemplo 1*) Un dado se lanza ocho veces; la probabilidad de obtener un 5 y un 6 dos veces es :

$$P_M = \frac{8!}{2!\ 2!\ 1!\ 1!\ 1!\ 1!}\ (1/6)^2\ (1/6)^2(1/6)(1/6)(1/6)(1/6) = 0{,}006$$

*Ejemplo 2*) En una urna se tienen 5 bolillas rojas, 4 blancas y 3 azules. Se saca una bola al azar de la urna y se anota su color, se la devuelve antes de hacer la siguiente extracción. Para calcular la probabilidad que de 6 pruebas realizadas, 3 sean rojas, 2 blancas y 1 azul

$$P_M = \frac{6!}{3!\ 2!\ 1!} (5/12)^3 (4/12)^2 (3/12)^1 = 0{,}12$$

*Ejemplo 3*) En una investigación epidemiológica retrospectiva respecto al cáncer de pulmón, se escogieron 75 pacientes de cáncer y 75 como control. Luego se averiguó entre ellos quienes fumaban y quienes no. Se encontraron los siguientes datos:

| | Enfermos | | |
|---|---|---|---|
| | SI | NO | Total |
| Fumador | 50 | 60 | 110 |
| No fumador | 25 | 15 | 40 |
| Total | 75 | 75 | 150 |

La probabilidad exacta de que ocurran esos sucesos se calcula con la Probabilidad Multinomial

$$P_M = \frac{150\,!}{50\,!\ 25\,!\ 60\,!\ 15\,!} (50/150)^{50}(25/150)^{25} (60/150)^{60} (15/150)^{15} = 0{,}00072$$

Odds de enfermedad: p / (1-p) = (75/150) / [1 – (75/150)] = TE / TS = 75 / 75 = 1
Odds del factor de riesgo: (110/150) / [( 1 – (110/150)] = 0,73 / 0,27 = 2,7 ≅ 3
Riesgo Relativo: RR = [(50/110)] / [(25/40)] = 0,73

Eso se interpreta como que hay 73 chances en 100 de contraer cáncer si se fuma.

# 7.7 Procesos de tipo Poisson

Se denominan procesos de tipo Poisson, o poissonianos, a todo experimento consistente en una serie de pruebas repetidas dentro de un *continuo*, caracterizadas por tener resultados que se pueden clasificar en si verifican o no, cierta propiedad o atributo, siendo aleatorios e independientes del lugar que ocurren dentro del continuo.

Para identificar un proceso Poisson en una serie de pruebas repetidas, se deben verificar tres condiciones:

1) *Sucesos puntuales:* Los sucesos ocurren dentro de un continuo (espacio o tiempo) y ocupan una parte infinitesimal del mismo. Es decir, en el espacio un suceso es puntual y en el tiempo es instantáneo. En términos prácticos, los sucesos no ocupan una parte apreciable del continuo.

2) *Sucesos independientes*: La ocurrencia de un suceso en un lugar del continuo no condiciona la ocurrencia del anterior (o del siguiente) en otra parte del mismo. Es decir, no hay reacción en ca-

dena, ni en cascada, ni hay apilamientos de células, ni agrupamientos en colonias, etc. Los sucesos son independientes del lugar que ocupan en el continuo.

3) *Probabilidad constante*: La probabilidad de ocurrencia de un suceso en un lugar del continuo es la misma en todo punto del mismo. Es decir, no hay lugares donde los sucesos ocurran con mayor frecuencia que en otros. Los puntos del continuo son equiprobables.

Son ejemplos de este tipo de proceso, un recuento celular en cámara de Neubauer o Hemocitómetro, la llegada de pacientes a una cola o línea de espera, los accidentes en una ruta, el desgaste de una pieza en una máquina de producción produce una reacción en cascada de productos defectuosos, y muchos otros casos. Esta probabilidad se *aproxima* a la binomial cuando la probabilidad de éxito es muy pequeña, por eso muchos la llaman: la ”binomial de los sucesos raros”.

Para imaginar casos de la vida real donde se la puede encontrar, se debe considerar un medio continuo como la sangre, orina, agua de mar o de río, el aire, el tiempo, hasta pasta de hornear de tamaño muy grande, dentro del cual ocurre un número total elevado de pequeñas cantidades discretas tales como los glóbulos blancos, rojos o plaquetas en la sangre, residuos en orina, plancton en el agua de mar, partículas de tierra en la de río, emisión de una partícula en sales de uranio, ralladura de limón en la pasta de una torta, etc. Si nada hace que tales partículas se agrupen (como la presencia de un coagulante), se cumple la condición de densidad uniforme de las partículas en el medio (tipo coloide). En caso contrario, se debe agitar o mezclar bien para lograrlo, tal como ocurre naturalmente en el torrente sanguíneo o en las aguas de un río, y artificialmente en la amasadora de una panadería, el vibrador de los laboratorios o el agregado de EDTA para evitar la coagulación en las muestras de sangre de los pacientes.

Otro modelo que produce casos Poisson son los que ocurren en el tiempo, como las llamadas que entran a una central telefónica, emisión de partículas radioactivas, pacientes o clientes en general esperando ser atendidos, autos que llegan a una estación de combustible, etc. En estos casos, se debe suponer: (a) los sucesos ocurren en forma independiente; (b) la probabilidad que un suceso ocurra en un corto intervalo sea proporcional a la longitud del intervalo; y (c) la duración del suceso es tan pequeña que hace insignificante a su probabilidad de ocurrencia. Entonces, si se cumplen estos tres supuestos la probabilidad de que ocurran exactamente r sucesos en un intervalo de tiempo finito es la de Poisson. En estos principios se basaron los Curie para su ley de Radiactividad Natural, que es la integral del valor medio de la función de Poisson variando en función del tiempo.

En términos generales, cuando un proceso del tipo Poisson tiene una “intensidad” promedio $\mu$ (es una especie de densidad de las partículas en el medio continuo), la probabilidad de que ocurran exactamente r sucesos o éxitos, se obtiene con:

$$\boxed{P_{PO}(r) = (e^{-\mu} \cdot \mu^{r}) / r!}$$

*Un proceso es de tipo Poisson, cuando los sucesos puntuales se producen individual y colectivamente al azar dentro de un continuo.*

# 7.7.1 Aplicaciones del Modelo Poisson

El primer ejemplo fue tomado del clásico estudio de Student sobre los recuentos celulares y publicado en la revista Biometrika en 1907. Se usa una cámara de recuento llamada en aquel entonces hemocitómetro, hoy más conocida como cámara de Neubauer. Para el recuento de glóbulos rojos se emplea la zona dividida en 400 cuadrados, como se conoce la cantidad de sangre y el grado de dilución, una vez contados los glóbulos se obtiene la concentración de los mismos por ml de sangre. Student (F. Gosset) la usaba para el conteo de células de levaduras de cerveza y observó al aplicar la teoría de errores gaussiana, que la función Poisson se adaptaba mejor que la de Gauss a los valores observados. Las implicancias históricas de este trabajo son grandes, pero en el ejemplo aquí desarrollado solo interesa ver la manera de calcular las frecuencias esperadas. Si se conoce la probabilidad de encontrar una cuadrícula vacía, mediante la fórmula Poisson se puede sacar el número esperado de cuadrículas vacías al multiplicarla por el número total de cuadrículas: $O(r=0) = N \cdot P_{PO}(r=0)$. En este caso se muestra la *manera iterativa* de resolver este tipo de problemas. La probabilidad Poisson para r sucesos es igual a la probabilidad anterior (r-1) multiplicada por $\mu$ y dividida por r . En su famoso informe, Student (1907) mostró que el ajuste usando Poisson era mejor que el usado hasta entonces, cosa que se verá más adelante en el tema Bondad de Ajuste estadística.

*Ejemplo 1) Recuentos celulares.*
Se trata de un recuento de células de levadura de cerveza en las 400 cuadrículas de un hemocitómetro. (Cuadro 4.2: Caso 3). La media aritmética allí calculada es una especie de densidad o concentración de células de levadura por unidad de volumen de agua. Es la intensidad en la fórmula de Poisson, o sea : $\mu = 0{,}6825$ células por cuadrícula. Entonces:

r = 0, 1, ...5 son los valores esperados de la cantidad de células por cuadrícula, se puede aplicar la fórmula y calcular las probabilidades para cada caso.

| Nº de células /cuadrícula ( r ) : | 0 | 1 | 2 | 3 | 4 | 5 | Σ |
|---|---|---|---|---|---|---|---|
| Frecuencias Observadas ( Oi ) : | 213 | 128 | 37 | 18 | 3 | 1 | 400 |
| Probabilidad Poisson : $P_{PO}(r)$ : | 0,505 | 0,345 | 0,118 | 0,027 | 0,0045 | 0,0005 | 1,00 |
| Frecuencias Esperadas ( Ei ) : | 202 | 138 | 47,2 | 10,8 | 1,8 | 0,2 | 400 |

Paso 1 : Se calcula la probabilidad de cuadrícula vacía:

$$P_{PO}(r=0) = e^{-\mu} = e^{-0{,}6825} = 0{,}505$$

Luego la frecuencia esperada será :

$$E(r=0) = N \cdot P_{PO}(r=0) = 400 \cdot 0{,}505 = 202$$

Paso 2 : Se calcula la probabilidad de encontrar una célula por cuadrícula

$$P_{PO}(r=1) = (e^{-\mu} \cdot \mu^1) / 1! = P_{PO}(r=0) \cdot \mu^1 / 1! = 0{,}505 \cdot 0{,}6825 = 0{,}345$$

Y así, es $E(r=1) = N \cdot P_{PO}(r=1) = 400 \cdot 0{,}345 = 138$

Paso 3 : Se calcula la probabilidad de encontrar dos células por cuadrícula

$$P_{PO}(r=2) = (e^{-\mu} \cdot \mu^2) / 2! = P_{PO}(r=1) \cdot \mu / 2 = 0{,}345 \cdot 0{,}6825 / 2 = 0{,}118$$

Y así, es $E(r=2) = N \cdot P_{PO}(r=2) = 400 \cdot 0{,}118 = 47{,}2$

Paso 4 : Se calcula la probabilidad de encontrar tres células por cuadrícula

$P_{PO}( r=3 ) = ( e^{-\mu} \ . \ \mu^3) / \ 3 ! \ = P_{PO}( r=2) \ . \ \mu \ / 3 = 0{,}118 \ . \ 0{,}6825 / 3 = 0{,}027$
Y así, es $E ( r=3 ) = N \ . \ P_{PO}( r=3 ) = 400 \ . \ 0{,}027 = 10{,}8$

Paso 5 : Se calcula la probabilidad de encontrar cuatro células por cuadrícula
$P_{PO}( r=4 ) = ( e^{-\mu} \ . \ \mu^4 ) / \ 4! \ = P_{PO}( r=3) \ . \ \mu / 4 \ = 0{,}027 \ . \ 0{,}6825 /4 = 0{,}0045$
Y así, es $E ( r=4 ) = N \ . \ P_{PO}( r=4 ) = 400 \ . \ 0{,}0045 = 1{,}8$

Paso 6 : Se calcula la probabilidad de encontrar cinco células por cuadrícula
$P_{PO}( r=5) = ( e^{-\mu} \ . \ \mu^5 ) / \ 5! \ = P_{PO}( r=4) \ . \ \mu / 5 \ = 0{,}0055 \ . \ 0{,}6825 / 5 \ = 0{,}0005$
Y así, es $E ( r=5 ) = N \ . \ P_{PO}( r=5 ) = 400 \ . \ 0{,}0005 = 0{,}2$

De la comparación entre las frecuencias observadas (Oi) y las esperadas (Ei), se puede deducir que tan bien se ajusta la teoría (Ei) a la realidad (Oi). Student demostró que si las Ei se las obtenía con el modelo de Poisson en lugar del gaussiano tradicional, el ajuste era mejor.

En el segundo ejemplo se toma un conteo radiológico de 100 cuentas de 0,1 minuto de una sola fuente, y también se puede apreciar un buen ajuste entre valores esperados y observados. Los datos reales se tomaron de la obra de Remington-Schork (ver Bibliografía).

*Ejemplo 2) Conteos radiológicos.*
Las frecuencias observadas de 100 cuentas radiológicas de 0,1 minuto, fueron medidas por J. Nehemías en un trabajo publicado en la obra antes citada.

| Conteo radiológico ( r ) : | 0 | 1 | 2 | 3 | 4 | 5 | 6 y más | $\Sigma$ |
|---|---|---|---|---|---|---|---|---|
| Frec. Observadas ( Oi ) : | 11 | 20 | 28 | 24 | 12 | 5 | 0 | 100 |
| Probab. Poisson $P_{PO}( r )$ : | 0,111 | 0,244 | 0,268 | 0,197 | 0,108 | 0,048 | 0,004 | 1 |
| Frec. Esperadas ( Ei ) : | 11,1 | 24,4 | 26,8 | 19,7 | 10,8 | 4,8 | 2,4 | 100 |

Repitiendo los pasos del ejemplo anterior, se vuelve a aplicar la manera iterativa para resolver estos casos.

Se ha comentado que la función de Poisson se ha utilizado reiteradamente para las teorías de fenómenos raros, tanto como para aplicaciones más prácticas. Para ilustrar estos puntos se dan a continuación dos ejemplos más:

*Ejemplo 3) Teoría de extinción de los dinosaurios.*
Budyko en 1974, usó la fórmula de Poisson para responder una pregunta clásica :
¿ La extinción de los dinosaurios en el Cretáceo fue por causas naturales o por algunas causas de tipo ambiental ?

Al final del período Cretáceo, de las diez órdenes de reptiles existentes cinco desaparecieron, a saber: dos órdenes de dinosaurios terrestres, una de dinosaurios de tipo volador, los pterosaurios y dos órdenes de reptiles acuáticos: el ictosaurio y el pleirosaurio. Estas cinco órdenes, incluyen treinta y cinco familias cada una de las cuales a su vez, estaba compuesta por una numerosa cantidad de especies. Se ha estimado el tiempo promedio de existencia de un orden de reptiles en el planeta, en 100 millones de años, y se ha calculado que la extinción se produjo en un período de 5 millones de años, al finalizar el Cretáceo.

Si se supone que la desaparición de un orden de dinosaurios es independiente de las otras, y que no hubo ningún desastre que haya desencadenado el proceso, se puede concluir que: "las extinciones se produjeron individual y colectivamente al azar". Entonces, el número promedio de extinciones será de $\mu$ = 5 millones de años / 100 millones de años = 0,05. Por lo tanto, la probabilidad que se produzcan 5 extinciones en 5 millones de años será de acuerdo a la fórmula de Poisson:

$$P_{PO}(\ r = 5\ ) = (\ e^{-0,05}\ .\ 0,05^{5}\ ) / 5! = 2,5\ .\ 10^{-9}$$

Un valor demasiado bajo como para suponer que se debió a causas naturales. Mas bien se puede pensar que una serie de circunstancias geológicas y climatológicas, unidas a grandes desequilibrios ecológicos, fueron determinantes para la desaparición de los grandes reptiles.

*Ejemplo 4*) *Técnica rápida de recuento en hemocitómetro.*
El efectuar un recuento de glóbulos rojos en cámara es una tarea larga y puntillosa. Es fácil equivocarse y por seguridad conviene hacerla más de una vez. Como demanda mucho tiempo y los honorarios son muy bajos, comparado con otras, esta técnica se hizo impopular. En su lugar, ahora se usa la técnica del microhematocrito (con una inversión baja), o mejor aún las cámaras de recuento automático que son aparatos de alto costo pero de alta producción. Sin embargo, hay lugares y situaciones donde se deben hacer las determinaciones y no se cuenta con los medios suficientes para tales aparatos mecánicos. Para esos casos especiales, se presenta esta técnica de conteo rápido. Aquí, lo que se deben contar no son las células sino las cuadrículas vacías en toda la cámara.

Sea una cámara típica de Neubauer, con una concavidad de 0,1 mm de profundidad y de 1 $mm^2$ de superficie, sobre un porta objeto, la misma está dividida en 20 x 20 cuadrículas; entonces el volumen de líquido sobre ellas será de: 0,1 mm . 1 $mm^2$ / 400 = (1/4000) $mm^3$. Se toma una muestra de sangre y se diluye por ejemplo al 25%, se homogeniza y se agita para evitar el efecto de apilamiento celular (o se agrega EDTA como anticoagulante). De esta forma se puede aceptar que los hematíes quedan distribuidos individual y colectivamente al azar dentro de la cuadrícula. Se dan las condiciones de un proceso del tipo Poisson. Sea H el número de glóbulos rojos en una cuadrícula, los que siguen una distribución en el continuo de tipo Poisson, con una intensidad $\mu$ (concentración de hematíes por unidad de volumen) desconocida y objeto de la medición. Entonces, la probabilidad de encontrar un número r de hematíes en una cuadrícula se obtiene con :

$$P_{PO}(\ H = r\ ) = (\ e^{-\mu}\ .\ \mu^{r}\ ) /\ r!$$

En particular si r = 0

$$P_{PO}(\ H = 0\ ) = e^{-\mu}$$

De donde

$$-\ln \{\ P_{PO}(\ H = 0\ )\ \} = \mu$$

Y para estimar la probabilidad de encontrar una cuadrícula vacía se pueden usar las frecuencias observadas. Basta contar la cantidad de cuadrículas vacías V y dividirla por el número total de cuadrículas N = 400. O sea,

$$P_{PO}(\ H = 0\ ) \approx V / N$$

Reemplazando es :

$$- \ln (V/400) = \mu = \ln ( 400 / V )$$

Como se había hecho una dilución de 400 volúmenes ( al 25% ), para obtener el número de glóbulos rojos por $mm^3$ : R , se deberá multiplicar el valor promedio hallado para una cuadrícula por un factor : f . Donde f es el volumen de líquido por cuadrícula ( 1/ 4000 $mm^3$ ) multiplicado por el factor de dilución (1/400). Esto es: f = 1 / 1.600.000

Entonces resultará :

$$f . R = \mu = \ln ( 400 / V )$$

De donde :

$$R = 1{,}6\ 10^6 . \ln ( 400 / V )$$

Por ejemplo, si se detectaron 20 cuadrículas vacías al hacer la medición el número de glóbulos rojos por $mm^3$ de sangre será: $R = 1{,}6\ 10^6 . \ln ( 400 / 40 ) = 3{,}68\ 10^6$ hematíes /$mm^3$

# 7.8 Aproximación de la binomial a la Poisson

Una binomial de muy baja probabilidad de éxito p y con un gran número de pruebas repetidas tales que su producto es finito, resulta ser muy parecida a una Poisson.

Si $p \to 0$ y $n \to \infty$ tales que $n.p \to \mu$ , entonces $P_B ( r/n,p ) \to P_{PO}( r / \mu )$

En la práctica, muchas veces aparecen casos con un número de pruebas lo suficientemente grande, como para ser considerado infinito. Con $n > 30$ y $p < 0{,}1$ el error que se comete al aproximar la binomial con la Poisson no tiene importancia. La ventaja, es que resulta mucho más cómodo calcular los valores de $P_{PO}$ que los de la $P_B$.

*Ejemplo 1*) Se sabe que el 1% de las jeringas descartables fabricadas tienen defectos en el cierre hermético. Se compra un lote y se prueban 30 de ellas. ¿Cuánto vale la probabilidad de que dos o más de ellas tengan defectos?

$$P_B ( r \geq 2/n=30,p=0{,}01) = P_B(r=2) + P_B(r=3) + ... + P_B(r=30) = 0{,}0328 + 0{,}031+...+0 = 0{,}0361$$
$$= 1 - P_B(r=0) - P_B(r=1) = 1 - 0{,}7397 - 0{,}2242 = 0{,}0361$$

$$P_{PO}( r \geq 2/\mu=0{,}3)=P_{PO}(r =2)+ P_{PO}(r =3)+...+ P_{PO}(r =30) = 0{,}0333 + 0{,}0033 + 0{,}0002 = 0{,}0368$$
$$= 1 - P_{PO}(r=0) - P_{PO}(r=1) = 1 - 0{,}7408 - 0{,}2222 = 0{,}037$$

Esta aproximación puede llegar a ahorrar mucho trabajo y tiempo en los cálculos. En el ejemplo anterior se puede apreciar que es bastante aceptable usar la Poisson, para calcular una Binomial, con importantes ventajas en cuanto a la simplicidad. A pesar de que estas aproximaciones dejaron de tener tanta importancia como antaño cuando no se disponía de computadoras para los cálculos, sigue siendo interesante ver el comportamiento de estas funciones de probabilidad en condiciones de borde para entender el hecho fundamental que se verá en el Teorema Central del Límite más adelante.

**J.C. Azzimonti Renzo: arroi_pss@ciudad.com.ar**

# 7.9 Aproximación de binomial a hipergeométrica

Cuando se cumplan las condiciones (1) o (2) siguientes:

(1) $r \ll R$ y $n \ll N-R$
o Entonces $P_B(r / n, p=R/N) \rightarrow P_H(r / n, N, R)$
(2) $n \ll R$ y $n-r \ll N-R$

*Ejemplo 1*) De 20 kits de glucosa guardados en el depósito, 8 de ellos tienen 5 o más semanas dentro del mismo y los 17 restantes son nuevos. Si se toman 4 kits al azar ¿ Cuánto vale la probabilidad que no más de tres de ellos sean viejos ?. Acá es N=20, R=8, r=2 y n=4.

$P_H(r \leq 2 / 4,20,8) = 1 - P_H(0 / 4,20,8) - P_H(1 / 4,20,8) = 1 - 0,1532 = 0,8468$

Pero como $r = 2 \ll R = 8$ y $8-4 = 4 \ll 12 = 20 - 8$ se puede aplicar la aproximación de la binomial con :

$P_B(r \leq 2 / 4 ; 0,4) = 1 - P_B(1 / 4 ; 0,4) - P_B(2 / 4 ; 0,4) = 1 - 0,1792 = 0,8208$

Se cometió un error de estimación del orden del 3% aproximadamente.

# 7.10 Problemas propuestos

1) Para identificar un proceso del tipo Bernoulli se deben verificar 3 condiciones. **V** **F**
2) Cualquier variable puede ser dicotomizada. **V** **F**
3) Si el resultado de una prueba incide en la siguiente se dice que son independientes. **V** **F**
4) La independencia de las pruebas es un supuesto básico de la teoría mendeliana. **V** **F**
5) La probabilidad de éxito se mantiene constante en un proceso Bernoulli. **V** **F**
6) Si los fenómenos ocurren en cascada, o en cadena, no es un caso de Bernoulli. **V** **F**
7) La probabilidad Binomial solo sirve para calcular la probabilidad de tener r éxitos. **V** **F**
8) La probabilidad Binomial se puede aplicar en Genética y en los nacimientos. **V** **F**
9) En un contagio la probabilidad de éxito permanece constante. **V** **F**
10) En un caso de repulsión la probabilidad de fracaso es una constante. **V** **F**
11) Las Probabilidades Binomial y Pascal están relacionas con una fórmula. **V** **F**
12) En Pascal se considera que se necesitan n pruebas para tener el último éxito. **V** **F**
13) Si se trata de las pruebas adicionales, para obtener r éxitos, es una hipergeométrica. **V** **F**
14) La Multinomial es una generalización de la Pascal. **V** **F**
15) Los sucesos puntuales ocurren individual y colectivamente al azar en una Poisson. **V** **F**
16) Si hay apilamiento celular en un recuento se siguen cumpliendo los supuestos Poisson. **V** **F**
17) En un modelo Poisson los sucesos ocurren en cadena o se agrupan en colonias. **V** **F**
18) En un modelo Poisson los sucesos ocurren en un continuo. **V** **F**
19) Los supuestos básicos del Modelo Poisson, son ................. ............................................... ...........
20) Los supuestos básicos del Modelo Bernoulli, son ............................................................... ...........
21) Un recuento celular es un ejemplo de aplicación del Modelo .......................
22) La probabilidad Binomial se puede aproximar con un Modelo .....................
23) La probabilidad Binomial se aproxima a una Hipergeométrica, bajo ciertas condiciones. **V** **F**
24) A la probabilidad Poisson se la llama la binomial de los casos raros. **V** **F**

**J.C. Azzimonti Renzo: arroi_pss@ciudad.com.ar**

2) En un apunte de cátedra de 200 páginas hay 220 errores de tipeo distribuidos al azar a lo largo del texto. Se pide calcular la probabilidad de que una página cualquiera tenga:

a) Ningún error. b) Dos errores. c) Dos o más errores.

3) En una fábrica de medicamentos el dosificador de droga en polvo produce un 1% de remedios defectuosos por mal funcionamiento. Si el farmacéutico a cargo del sector Control de Calidad elige al azar 100 medicamentos de un lote de producción, se pide calcular la probabilidad de encontrar: a) Ningún defectuoso. b) Exactamente 2 defectuosos. c) Más de 3 defectuosos.

4) En la población de alumnos de la Facultad de Ciencias Exactas hay un 5% de zurdos. Si se eligen al azar 100 estudiantes calcular la probabilidad de que: a) 97 o más usen su mano derecha. b) 2 o más sean zurdos. c) Ninguno sea zurdo. d) Todos usen la mano izquierda.

5) Si en la población de la Provincia de Misiones estimada en 800.000 personas ocurren en promedio 80 suicidios al año, calcular la probabilidad de que el año que viene ocurran: a) 90 suicidios. b) 80 o más suicidios. c) Menos de 80 suicidios.

6) De un mazo de barajas españolas de 40 cartas se sacan 3 cartas, hallar la probabilidad de que: a) Sean las 3 del mismo palo si se reponen a medida que se sacan. b) Ídem sin reposición. c) Por lo menos una sea de oros. d) Dos sean de copas. e) Ninguna sea de espadas. f) Las 3 sean figuras. g) Salgan 3 ases. h) Sacar el 7 de oros, el 1 de espadas y el 1 de bastos.

7) En la cátedra de Bioestadística hay 8 auxiliares que pueden trabajar por igual en su casa tanto como en la Facultad. Si todos se destinan a un mismo gabinete, se desea averiguar cuantos escritorios hay que colocar para que cada uno tenga en el 90% de las veces un escritorio disponible.

8) En una farmacia se reciben en promedio 5 pedidos por TE para entrega a domicilio por día. En una hora cualquiera elegida al azar, calcular la probabilidad de que: a) Se reciban 3 pedidos. b)Se reciban más de 5 pedidos. c) Se reciban entre 2 y 4 pedidos.

9) En un hospital se mide cuantos enfermos con infecciones venéreas solicitan atención en un período de 50 días. Los datos se muestran en la tabla siguiente. En la primer columna se colocan las frecuencias observadas diarias. En la segunda la cantidad de días en que ocurrieron. Y en la última columna se calcularon las frecuencias relativas (probabilidades empíricas) observadas. Se pide calcular en una cuarta columna las probabilidades teóricas respectivas usando un modelo de Poisson con una media de 5,66 pacientes diarios.

| **Pacientes diarios** | **Nº de días** | **Frec. Relativa Empírica** |
|---|---|---|
| 3 | 3 | 0,06 |
| 4 | 7 | 0,14 |
| 5 | 12 | 0,24 |
| 6 | 14 | 0,28 |
| 7 | 10 | 0,20 |
| 8 | 4 | 0,08 |
| Total | 50 | 1,00 |