Universidad Nacional del Litoral

Facultad de Ingeniería y Ciencias Hídricas

ESTADÍSTICA

Ingenierías: Recursos Hídricos-Ambiental-Agrimensura-Informática

Mg. Ing. Susana Vanlesberg

Profesor Titular

MODELOS DE VARIABLE ALEATORIA DISCRETA

INTRODUCCIÓN

- La aplicación de la teoría de probabilidad en situaciones reales, concretas, que presentan una regularidad en el planteo, ha originado una serie de modelos que las resuelven.

- Los ingenieros realizan supuestos respecto del problema a resolver, esto los lleva a descripciones análogas y a formas matemáticamente iguales a las de los modelos probabilísticos.

En este capítulo no sólo se presentan los modelos para variables aleatorias discretas más usados para fines prácticos, sino que también se brindan algunas nociones de los mecanismos por los cuales se ha originado cada uno. Esto es de suma importancia para un ingeniero, ya que la existencia de tales mecanismos puede describir una situación física de su interés, y ésta es una razón más importante que la buena aproximación matemática de algún modelo.

- Modelo Bernoulli
- Modelo Binomial
- Modelo Geométrico
- Modelo Poisson
- Modelo Hipergeométrico

- En cada modelo se va a determinar la función que permite calcular probabilidades en un valor y acumuladas, su esperanza y su varianza

Situaciones a resolver:

- 1- Estudios de medio ambiente realizados en la ciudad han evaluado que en el 30% de los hogares se separan los residuos orgánicos del resto.
- La Dirección de Limpieza de la Municipalidad selecciona un barrio al azar, dentro del barrio una manzana y, en la misma, elige 10 hogares en forma aleatoria.

- a) En base a estos datos, la Municipalidad quiere saber la probabilidad de que se separen los residuos en la mitad de los hogares de la ciudad.
- b) Un análisis más minucioso estableció que, de ese 30%, un 1% no realizaba una separación correcta.
- Si se obtiene una muestra de 500 hogares en toda la ciudad, ¿Cuál es la probabilidad de que en por lo menos 5 hayan realizado mal la separación?

- 2-Se cuenta con registros de cantidad de días con lluvia superior a 150mm en una determinada estación meteorológica.
- Se sospecha que entre los registros hay valores que son erróneos, el técnico que realiza las observaciones dice que en general el 80% son correctos, entonces se analiza la población de 100 valores.

- Se seleccionó una muestra aleatoria sin reemplazo de 6 valores.

- A partir de esta información se quiere saber que probabilidad hay de que entre los registros seleccionados haya al menos uno equivocado.

Modelo Bernoulli

- Los resultados de los experimentos pueden separarse en dos categorías mutuamente excluyentes: *éxito* o *fracaso*; por ejemplo mido y me equivoco o mido y no me equivoco, o llueve o no llueve etc.
- Puede entonces definirse la variable aleatoria x de tipo Bernoulli y asignársele valores (arbitrarios pero muy prácticos) a los eventos antes mencionados:

- x=0 fracaso
- x=1 éxito

$$f(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

- Esta función deberá cumplir con las condiciones de una función de probabilidad

Características

$$E(x) = \sum_{\forall x_i} x_i \; f(x_i) = 0.(1 - p) + 1.p = p$$

$$V(x) = \sum_{x_i} (x_i - E(x))^2 \; f(x_i) =$$

$$= (0 - p)^2 (1 - p) + (1 - p)^2 .p =$$

$$Var = p(1 - p)$$

Modelo Binomial

- Si se realizan una serie de pruebas de tipo Bernoulli cuyos resultados sean independientes y si la probabilidad de éxito permanece invariable en todas ellas se origina un nuevo modelo denominado BINOMIAL.

- La expresión de la función surge de considerar todas las combinaciones posibles de x éxitos en n pruebas con probabilidad del suceso en estudio igual a **p**

- Para generalizar se considera el caso de 3 pruebas:

- \- Ningún éxito: 0 ; 0 ; 0 ; **$(1 - p)^3$**
- \- Un éxito : 1 ; 0 ; 0 ; ó 0 ; 1 ; 0 ; ó 0 ; 0 ; 1
- cada secuencia es un evento excluyente su probabilidad es

$p\,(1 - p)^2$; luego:

$$\mathbf{3\,p\,(1 - p)^2}$$

- \- Dos éxitos :
- 1 ; 1 ; 0 ; ó 1 ; 0 ; 1 ; ó 0 ; 1 ; 1 ;
- nuevamente cada secuencia es un evento excluyente de los restantes $p^2\,(1 - p)$ luego
- **$3\,p^2\,(1 - p)$**
- \- Tres éxitos: 1 ; 1; 1; p ; p ; p **p^3**

- **generalizando**

$$P(X = x) = \binom{n}{x} p^{x} (1 - p)^{n-x}$$

$$coeficiente\ binomial \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

- La función de distribución:

$$F(x) = P(X \leq x) = \sum_{\forall x_i \leq x} f(x_i)$$

Características

$$E(X) = \sum_{x_i=0}^{n} x_i \, f(x_i) =$$

$$= 0.\frac{n!}{0!(n-0)!} p^0 q^n + 1.\frac{n!}{1!(n-1)!} p^1 q^{n-1} + \ldots + n.\frac{n!}{n!(n-n)!} p^n q^0 = = np(p+q)^{n-1}$$

$$E(x) = np$$

$$Var(X) = E[X - E(X)]^2 = E(X^2) - E^2(X)$$

$$E(X^2) = \sum_{x_i=0}^{n} \binom{n}{x_i} p^{x_i} q^{n-x_i} x_i^2 =$$

$$ya\ que\ el\ primer\ termino\ es\ cero\ ,$$

$$la\ suma\ se\ toma\ desde\ 1$$

$$= np\sum_{x_i=1}^{n}\frac{(n-1)!}{x_i!\,(n-x_i)!}\,p^{x_i-1}q^{n-x_i}\,x_i^2$$

$$= np\sum_{x_i=1}^{n}\frac{(n-1)!}{x_i(x_i-1)!\,(n-x_i)!}\,p^{x_i-1}q^{n-x_i}\,x_i^2$$

$$haciendo\; x_i = x_j + 1,\; para\; x_i = 1, x_j = 0, x_i = n, x_j = n-1$$

$$x_i! = (x_i-1)!\,x_i = x_j!\,(x_j+1)$$

$$np\sum_{x_j=0}^{n-1}\frac{(n-1)!}{(x_j)!\,(x_j+1)(n-x_j-1)!}\,p^{x_j}q^{n-1-x_j}(x_j+1) =$$

$$np\left[\sum_{x_j=1}^{n-1}\binom{n-1}{x_j}p^{x_j}q^{n-1-x_j}x_j + \sum_{x_j=0}^{n-1}\binom{n-1}{x_j}p^{x_j}q^{n-1-x_j}\right]$$

$$El\ 2do\ termino\ de\ la\ suma\ es\ 1;\ se\ analiza\ el\ 1^{o},$$

$$ya\ que \binom{n-1}{x_j} x_j = \frac{(n-1)!\,x_j}{x_j!\,(n-1-x_j)!} = \frac{(n-2)!\,(n-1)\,x_j}{(x_j-1)!\,x_j\,(n-2-(x_j-1))!}$$

$$sacando\ factor\ comun\ (n-1)p\ ,\ se\ consigue$$

$$\sum_{x_j=1}^{n-1} \binom{n-2}{x_j-1} p^{x_j-1} q^{n-1-x_j}$$

$$haciendo\ x_j = x_k + 1\ ,\ x_k\ varia\ de\ 0\ a\ n-2\ con\ lo\ cual$$

$$\sum_{x_k=0}^{n-2} \binom{n-2}{x_k} p^{x_k} q^{n-2-x_k} = 1$$

$$E(X^2] = np[(n-1)p+1] = n^2 p^2 - n p^2 + np$$

$$Var(X) = n^2p^2 - np^2 + np - n^2p^2 = np(1 - p)$$

$$Var(X) = npq$$

- Estos mismos resultados podrían haberse obtenido de forma mucho más sencilla por considerar a la variable aleatoria X como la suma de n variables independientes idénticamente distribuidas como Bernoulli, con esperanza **p** y varianza **p q**

Modelo Geométrico

- Pruebas repetidas, independientes, con dos posibles resultados: ocurrencia o no del evento en estudio, con probabilidad favorable p = constante.
- Se deriva el modelo correspondiente a la variable **N**: número de pruebas hasta que ocurre el primer éxito. Este primer éxito se producirá sí y sólo sí en las pruebas anteriores no se produjo, lo cual ocurre con probabilidad igual a **(1 - p)**

$$P(N=n)= f(n)=(1-p)^{n-1}\, p$$

$$con \quad n = 1,2,3.....$$

$$F(n)=\sum_{j=1}^{n} p(j)=\sum_{j=1}^{n}(1-p)^{j-1}\, p =$$

$$= p\left[\frac{(1-p)^{n}-1}{(1-p)-1}\right]=1-(1-p)^{n}$$

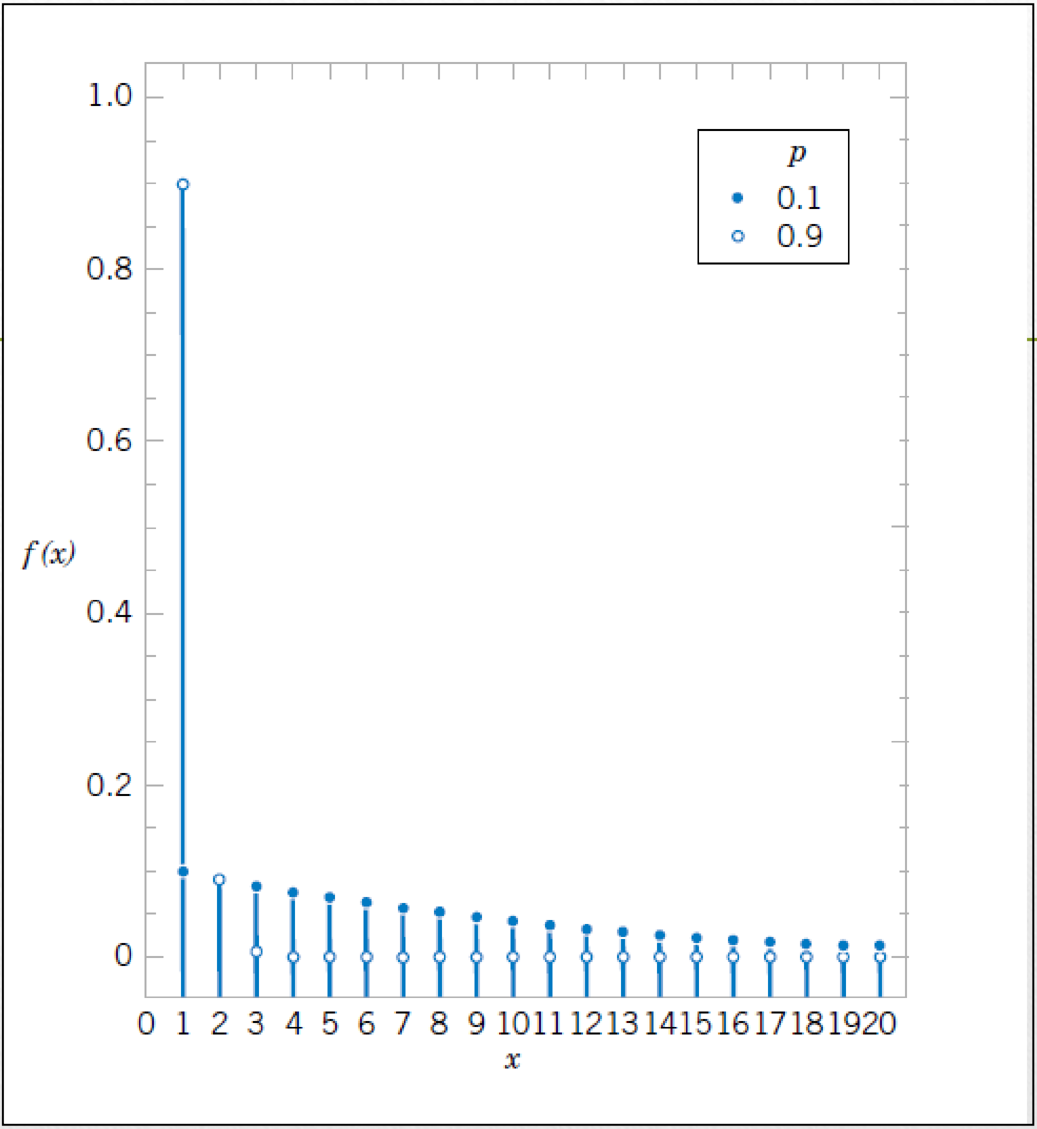
$$\left[\frac{(1-p)^{n}-1}{(1-p)-1}\right] desarrollo\ de\ la$$

$$suma\ de\ una\ progresión\ geométrica$$

- Esto mismo se podría haber obtenido al considerar la probabilidad de que N ≤ n simplemente como la probabilidad que exista al menos una ocurrencia en n pruebas:

$$P(N \geq 1) = 1 - P(N < 1) = 1 - P(\text{no haya ocurrencias en } n \text{ pruebas}) =$$

$$= 1 - (1 - p)^n$$



1.0
0.8
0.6
0.4
0.2
0
f(x)
p
0.1
0.9
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
x

Características

$$E(N) = \sum_{n=1}^{\infty} n\, p\, (1 - p)^{n-1} = p \sum_{n=1}^{\infty} n\, (1 - p)^{n-1}$$

$$E(N) = p \frac{1}{p^2} = \frac{1}{p}$$

- Este resultado significa que el número promedio de pruebas hasta que ocurre el primer éxito es la inversa de la probabilidad p de ocurrencia del evento.
- Esto se asocia al concepto de **período de retorno**, sumamente utilizado en ingeniería.

$$Var(N) = E(N^2) - E^2(N) = \left[\sum_{n=1}^{\infty} n^2 p(1-p)^{n-1}\right] - \left(\frac{1}{p}\right)^2 =$$

$$= \frac{1}{p^2}(1+q-1)$$

$$Var(N) = \frac{1-p}{p^2}$$

- **El término período de retorno es muchas veces mal interpretado.**
- **Si se dice que es 100 se cree que pasarán 100 años entre ocurrencias de valores de una variable de interés y superiores a un valor crítico; pero esto no es así, ya que la probabilidad de que se produzca un valor mayor al crítico permanece invariable para cualquier año, independientemente de lo que ocurrió en años anteriores. Es el valor en promedio!!**

Modelo Poisson

- Si **n** se incrementa y **p** se hace pequeña, pero el número promedio de eventos en el intervalo total permanece constante e igual a **p . n el modelo Binomial da lugar al modelo de Poisson.**
- Considerando la función de probabilidad del modelo Binomial en el límite,
- $\mathbf{p \to 0 \ y \ n \to \infty \quad y \quad \lambda = n \cdot p}$

$$\lambda = n\,p \quad \rightarrow p = \frac{\lambda}{n}$$

$$f(x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} =$$

$$= \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \frac{n!}{(n-x)!n^x\left(1-\frac{\lambda}{n}\right)^x} \cong 1$$

$$=$$

$$f(x) = P(X = x) = \frac{e^{-\lambda}\,\lambda^x}{x!}$$

Características

$$E(X)=\sum_{x=0}^{\infty} x\,\frac{e^{-\lambda}\,\lambda^{x}}{x!}=$$

$$\lambda\sum_{x=1}^{\infty} x\,\frac{e^{-\lambda}\,\lambda^{x-1}}{x\,(x-1)!}$$

$$E(X)=\lambda$$

$$\sum_{x=1}^{\infty}\frac{e^{-\lambda}\,\lambda^{x-1}}{(x-1)!}=1$$

$$Var(X) = E(X^2) - E^2(X)$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!} = \lambda \left[\sum_{x=0}^{\infty} \frac{x^2 \lambda^{x-1} e^{-\lambda}}{x(x-1)!} \right] =$$

$$Var(X) = \lambda$$

- Generalmente este modelo se vincula a aquellos eventos que ocurren en una unidad de tiempo, luego el período de tiempo en el que se realiza el análisis constituye una secuencia de pruebas independientes cada una con distribución Binomial.

- Si se tomara para el análisis un intervalo de tiempo igual al doble o al triple del inicial el parámetro es también igual al doble, al triple, etc., marcando esto la dependencia del tiempo de este modelo y por ello vinculado a los procesos estocásticos.
- Se entiende por procesos estocásticos a aquellos en los que interesa la secuencia, en el tiempo, de ocurrencia de eventos.

- Un fenómeno para ser clasificado como proceso de Poisson debe cumplir con las siguientes condiciones:

- **a- Estacionariedad**: esto significa que la probabilidad de que ocurra un evento dado en un intervalo de tiempo muy pequeño Δt, es proporcional a ese tiempo e igual a $\lambda\Delta t$.

- **b- No multiplicidad**: esto es que la probabilidad de que ocurran dos o más eventos en un intervalo de tiempo Δt es despreciable comparado con $\lambda\Delta t$.

- **c- Independencia**: el número de eventos en algún intervalo de tiempo es independiente del número de eventos en algún otro intervalo de tiempo.

Modelo Hipergeométrico

- Este modelo surge cuando se realiza un **muestreo sin reposición** de una **población finita** con sus elementos clasificados en dos categorías.
- Si **N** es el total de elementos de los cuales hay **k** de una categoría y **N-k** de otra, al realizar una extracción de **n** elementos, sin reposición, cada extracción que se realice posteriormente es dependiente del resultado de la extracción anterior con lo cual **va cambiando la probabilidad de éxito**.

$$P(X = x) = f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Características

- Para obtener las características es posible decir que la variable aleatoria X es la suma de n variables xi como en el caso Binomial, pero con la diferencia que aquí las xi son dependientes.
- Como para sumar las esperanzas no se necesita que las variables aleatorias sean independientes entonces:

$$E(X) = E(x_1) + E(x_2) + \ldots\ldots + E(x_n)$$

- Cada E (xi) es la probabilidad de x en la i ésima prueba: k/N , si no se sabe que ha ocurrido en pruebas anteriores o posteriores:

$$E(X) = n\left(\frac{k}{N}\right)$$

- La varianza en cambio no es aditiva para variables dependientes. Pero se obtiene la siguiente expresión:

$$Var(X) = n * p * q * \left(\frac{N - n}{N - 1} \right)$$

(N-n/ N-1) es el factor de corrección por muestreo sin reposición.

Resolución de las Situaciones a resolver:

- Estudios de medio ambiente realizados en la ciudad han evaluado que en el 30% de los hogares se separan los residuos orgánicos del resto.
- La Dirección de Limpieza de la Municipalidad selecciona un barrio al azar, dentro del barrio una manzana y, en la misma, elige 10 hogares en forma aleatoria.

- a) En base a estos datos, la Municipalidad quiere saber la probabilidad de que se separen los residuos en la mitad de los hogares de la ciudad.

- b) Un análisis más minucioso estableció que, de ese 30%, un 1% no realizaba una separación correcta.
- Si se obtiene una muestra de 500 hogares en toda la ciudad, ¿Cuál es la probabilidad de que en por lo menos 5 hayan realizado mal la separación?

- En primer lugar, debemos definir de qué variable se trata, cómo está definida, y cuál es la distribución de probabilidades que se ajusta al planteo del problema.
- Se trata de una variable **discreta** (las cuales están asociadas a problemas de **conteo**:
- ***X = {Cantidad de hogares donde se realiza la separación de residuos}***

- El experimento es de tipo Bernoulli: tiene solamente dos resultados posibles:
    - ***Se separan los residuos.***
    - ***No se separan los residuos.***
- Al repetir el experimento ***n*** veces, se obtiene la ***DISTRIBUCIÓN BINOMIAL***

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

Donde P = probabilidad de que x se observe exactamente x veces

n = cantidad de ensayos = 10

p = probabilidad de éxito = 0.3

q = probabilidad de fracaso = 0.7

El cálculo de la probabilidad resulta entonces:

$$P(X = 5) = \binom{10}{5} \times 0.7^5 \times (1 - 0.7)^{10-5} \cong 0.1029$$

A este resultado se puede llegar de varias maneras.

1) Resolviendo paso a paso:

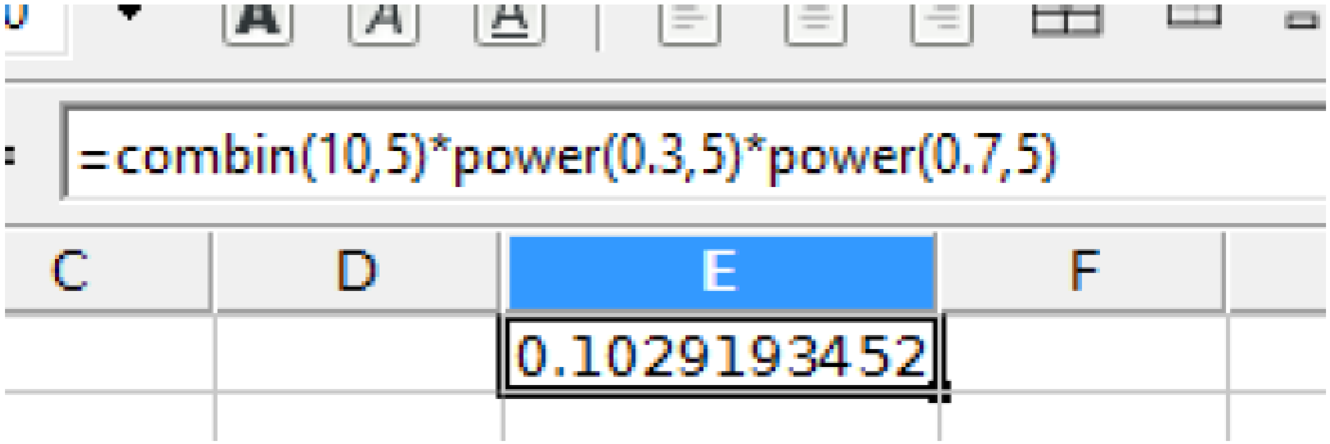
$$\binom{10}{5} = \frac{10!}{(10-5)! \times 5!} = 252 \qquad \text{(i)}$$

$$0.3^5 = 0.00243 \qquad \text{(ii)}$$

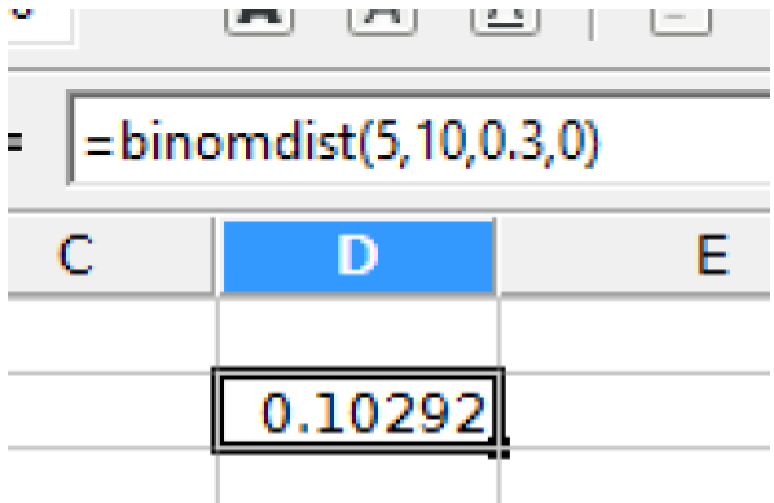
$$0.7^{10-5} = 0.16807 \qquad \text{(iii)}$$

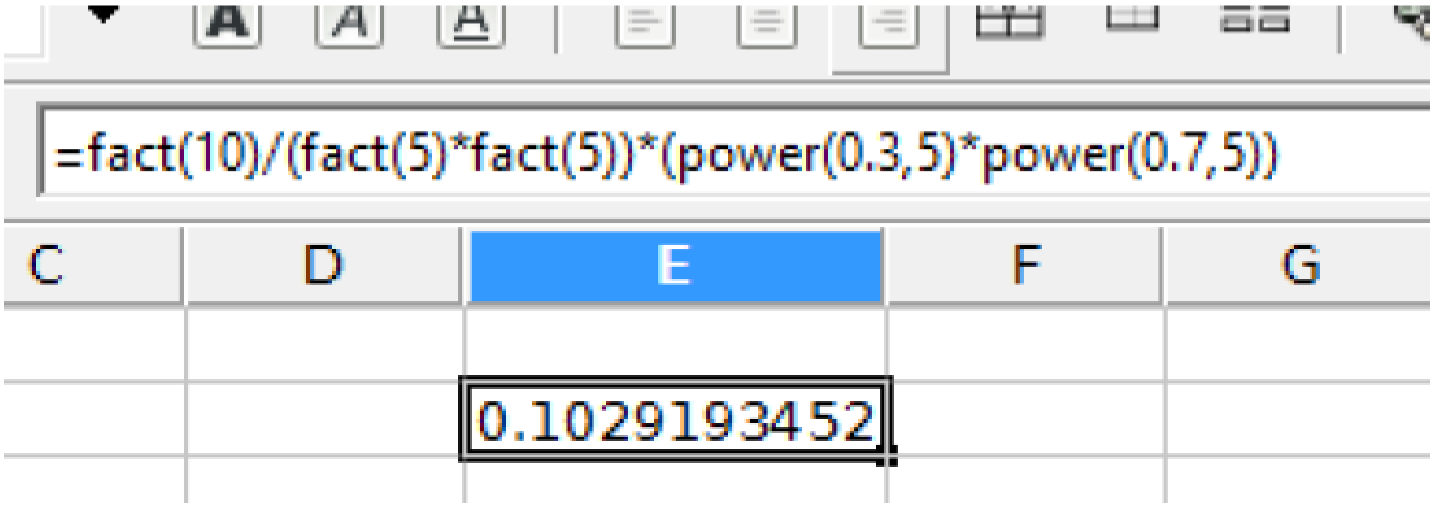
$$\text{(i)} \times \text{(ii)} \times \text{(iii)} = 0.10292$$

2) Empleando una planilla de cálculo (p.ej. Gnumeric):



O:




=fact(10)/(fact(5)*fact(5))*(power(0.3,5)*power(0.7,5))
C
D
E
F
G
0.1029193452

O realizando una búsqueda en tablas:

TABLA 5 (CONTINUACIÓN)

						p					
n	k	0.05	0.10	0.15	0.20	0.25	$\boxed{0.30}$	0.35	0.40	0.45	0.50
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
9	1	0.2985	0.3874	0.3679	0.3020	0.2253	0.1556	0.1004	0.0605	0.0339	0.0176
9	2	0.0629	0.1722	0.2597	0.3020	0.3003	0.2668	0.2162	0.1612	0.1110	0.0703
9	3	0.0077	0.0446	0.1069	0.1762	0.2336	0.2668	0.2716	0.2508	0.2119	0.1641
9	4	0.0006	0.0074	0.0283	0.0661	0.1168	0.1715	0.2194	0.2508	0.2600	0.2461
9	5	0.0000	0.0008	0.0050	0.0165	0.0389	0.0735	0.1181	0.1672	0.2128	0.2461
9	6	0.0000	0.0001	0.0006	0.0028	0.0087	0.0210	0.0424	0.0743	0.1160	0.1641
9	7	0.0000	0.0000	0.0000	0.0003	0.0012	0.0039	0.0098	0.0212	0.0407	0.0703
9	8	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0013	0.0035	0.0083	0.0176
9	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0020
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
10	1	0.3151	0.3874	0.3474	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098
10	2	0.0746	0.1937	0.2759	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439
10	3	0.0105	0.0574	0.1298	0.2013	0.2503	0.2668	0.2522	0.2150	0.1665	0.1172
10	4	0.0010	0.0112	0.0401	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051
$\boxed{10}$	$\boxed{5}$	0.0001	0.0015	0.0085	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461
10	6	0.0000	0.0001	0.0012	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051
10	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

En la segunda pregunta, el concepto es el mismo ***pero***:

I) La probabilidad es muy pequeña (tiende a cero)

$$p = 0.3 \times 0.01 = 0.003$$

II) n es muy grande (tiende a infinito), n = 500

Entonces nos encontramos con una ***distribución de Poisson:***

En esta distribución, se tiene que:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad \text{donde:}$$

$$x = \text{ cantidad de éxitos}$$

$$\lambda = np$$

$$P(X \leq x) = \sum_{i=0}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

En el caso que nos ocupa, nos interesa estimar la probabilidad de que se realice la separación de residuos en más de 5 hogares.

$$P(X \geq 5)$$

Esto es equivalente a que no se separen los residuos en 4 hogares o menos:

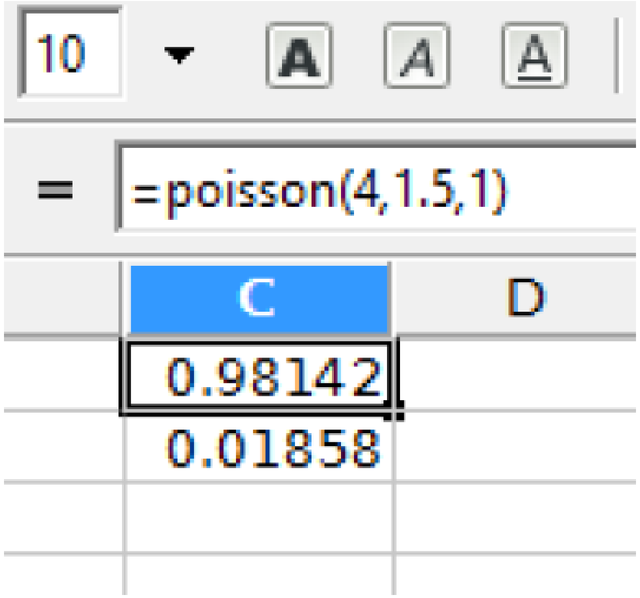
$$P(X \geq 5) = 1 - P(X < 5) = 1 - \sum_{x=0}^{4} \frac{e^{-500 \times 0.003}(500 \times 0.003)^{x}}{x!}$$

Esto también puede resolverse de varias maneras:

I) Extensivamente, usando una planilla de cálculo:

	A	B	C	D
1	lambda=nxp=		500x0.003=	
2				
3	x	exp(-lambda)*lambda**x/fact(x)		
4	0	0.22313016014843		
5	1	0.33469524022264		
6	2	0.25102143016698		
7	3	0.12551071508349		
8	4	0.04706651815631		
9	suma	0.98142406377786		
10	1-suma	0.01857593622214		
11				
12				

II) Empleando la fórmula en la planilla de cálculo:



III) Usando las tablas:

7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.

					λ		
k	1.1	1.2	1.3	1.4	1.5	1.6	
0	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.
1	0.3662	0.3614	0.3543	0.3452	0.3347	0.3230	0.
2	0.2014	0.2169	0.2303	0.2417	0.2510	0.2584	0.
3	0.0738	0.0867	0.0998	0.1128	0.1255	0.1378	0.
4	0.0203	0.0260	0.0324	0.0395	0.0471	0.0551	0.
5	0.0045	0.0062	0.0084	0.0111	0.0141	0.0176	0.
6	0.0008	0.0012	0.0018	0.0026	0.0035	0.0047	0.
7	0.0001	0.0002	0.0003	0.0005	0.0008	0.0011	0.

2-

- Resolvemos con el modelo Hipergeométrico ya que tenemos dos clases de registros: los erróneos y los no erróneos.

- Se hace un muestreo sin reposición de una población finita:

$$N = 100 \quad K = 20 \quad N - K = 80$$

$$n = 6$$

$$P(X \geq 1) = 1 - P(X < 1) =$$

$$= 1 - P(X = 0) =$$

$$= 1 - \frac{\binom{20}{0}\binom{80}{6}}{\binom{100}{6}} = 1 - 0.252 =$$

$$\cong 0.75$$