VECTORES ALEATORIOS

Hagamos en primer lugar algunas consideraciones respecto de β^k, tribu de Borel en R^k. Recordemos que la tribu de Borel de un espacio topológico viene engendrada por el sistema de abiertos (ó cerrados) del espacio en cuestión. No obstante, hay ocasiones en que, si el espacio lo permite, podemos fijar nuestra atención en subfamilias que engendran igualmente dicha tribu. Así por ejemplo, recordemos que en el caso de R, β podía venir engendrada por intervalos de la forma $]-\infty, a]$, $a \in R$ o bien por cualquier otra familia de intervalos. Igual ocurre en R^k, de manera que podemos engendrar β^k a partir de cualquier familia de intervalos, por ejemplo $]-\infty, \vec{x}]$ con $\vec{x} = (x_1, \ldots, x_k)$. Pasemos ahora a la definición de vector aleatorio.

<u>Definición</u>. Sea $(S, \mathcal{A}, P)$ un espacio de probabilidad y sea el espacio medible (R^k, β^k). Definimos entre ambos una aplicación medible $\vec{X}$, a la que denominaremos vector aleatorio.

El concepto de vector aleatorio no es más que una generalización inmediata del concepto de variable aleatoria que ya hemos estudiado. De la definición que acabamos de dar se deriva de inmediato la siguiente propiedad

<u>Propiedad</u>.- Una aplicación $\vec{X}$ de $(S, \mathcal{A}, P)$ en (R^k, β^k) es un vector aleatorio si y sólo si cada una de sus componentes X_i, $i = 1, \ldots, k$ es una variable aleatoria.

Demostración.- Sea $\vec{X}$ un vector aleatorio y definamos sobre R^k la función g_i de la forma

$$g_i(x_1 \ldots x_k) = x_i$$

g_i es la proyección y como tal es continua y por tanto medible, en consecuencia

$$X_i = g_i(\vec{X})$$

es una variable aleatoria por cuanto resulta de la composición de dos aplicaciones medibles.

Sean ahora las X_i, $i = 1, \ldots, k$, variables aleatorias. Para demostrar que $\vec{X}$ es un vector aleatorio hemos de comprobar que

$$\vec{X}^{-1}(B) \in \mathcal{A} \quad , \quad B \in \beta^k$$

pero teniendo en cuenta las consideraciones iniciales basta con tomar B en una de las familias que engendran β^k, por ejemplo $B =]-\infty, \vec{x}]$, entonces

$$\vec{X}^{-1}(B) = \{s \in S;\ \vec{X}(s) \in B\} = \{s \in S;\ X_i(s) \in]-\infty, x_i],\ i = 1, \ldots, k\} =$$

$$= \bigcap_{i=1}^{k} \{s \in S;\ X_i(s) \in]-\infty, x_i]\} = \bigcap_{i=1}^{k} \{X_i^{-1}(]-\infty, x_i])\} \in \mathcal{A}$$

y este último conjunto es de $\mathcal{A}$ por serlo cada $X_i^{-1}(]-\infty, x_i])$, ya que las X_i son variables aleatorias.

<u>Probabilidad inducida</u>. A partir de P y mediante $\vec{X}$ podemos definir sobre el espacio medible (R^k, β^k) una medida de probabilidad. En efecto, para cada $B \in \beta^k$

$$P'(B) = P(\vec{X}^{-1}(B)).$$

Análogamente a como lo hacíamos en el caso unidimensional podemos ahora definir una <u>Función de distribución de probabilidad</u> a partir de esta probabilidad inducida tomando una clase especial de conjuntos de Borel. En efecto, para $B =]-\infty, \vec{x}]$, tenemos

$$F(\vec{x}) = F(x_1, \ldots, x_k) = P'(]-\infty, \vec{x}]) = P(\vec{X}^{-1}(]-\infty, \vec{x}])) = P(\vec{X} \leq \vec{x}).$$

De esta definición se derivan, de inmediato, las <u>siguientes propiedades</u>:

1) La función $F(\vec{x})$ es continua por la derecha y no decreciente en cada una de las componentes de $\vec{x}$.
La comprobación es sencilla y similar a la presentada en el caso unidimensional.

2) Sean $a = (a_1, \ldots, a_k)$ y $b = (b_1, \ldots, b_k)$, tales que $a \leq b$, es decir $a_i \leq b_i$, $i = 1, \ldots, k$. El subconjunto de R^k, $\{a \leq b\}$ representa un rectángulo k-dimensional con extremos superior e inferior, b y a respectivamente.

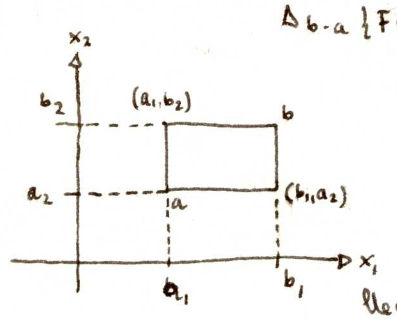
Definamos ahora el siguiente operador $\Delta_{b_i - a_i}$ que actuará sobre cualquier función definida en R^k de la siguiente forma

$$\Delta_{b_i - a_i}\{F(a_1, \ldots, a_i, \ldots a_k)\} = F(a_1, \ldots, b_i, \ldots a_k) - F(a_1, \ldots, a_i, \ldots, a_k).$$

Introducimos también el operador Δ_{b-a}, de la siguiente forma:

$$\Delta_{b-a}\{F(a)\} = \Delta_{b_1 - a_1} \Delta_{b_2 - a_2} \cdots \Delta_{b_k - a_k}\{F(a)\}$$

Así, por ejemplo, para $k = 2$, tendríamos



$$\Delta_{b-a}\{F(a)\} = \Delta_{b_1 - a_1} \cdot \Delta_{b_2 - a_2}\{F(a_1, a_2)\} = \Delta_{b_1 - a_1}\{F(a_1, b_2) - F(a_1, a_2)\}$$

$$= \Delta_{b_1 - a_1}\{F(a_1, b_2)\} - \Delta_{b_1 - a_1}\{F(a_1, a_2)\} = F(b_1, b_2) - F(a_1, b_2) - (F(b_1, a_2) - F(a_1, a_2)) = [F(b_1, b_2) - F(b_1, a_2)] - [F(a_1, b_2) - F(a_1, a_2)]$$

y observando la figura y teniendo en cuenta que $F(\vec{x}) = P(\vec{X} \leq \vec{x})$ llegamos a que $\Delta_{b-a}\{F(a)\} = P(\vec{X} \in]a, b]) \geq 0$

Esta es precisamente la segunda propiedad de $F(\vec{x})$, a saber:

Si $a \leq b$, entonces $\Delta_{b-a}\{F(a)\} \geq 0$.

3) $\lim_{x_i \to -\infty} F(x_1, x_2, \ldots, x_k) = 0$, para cada x_i.
Se comprueba fácilmente observando que los intervalos de la forma $\{\vec{t};\ \vec{t} \leq \vec{x}\}$ tienden al vacío a medida que $x_i \to -\infty$.

4) $\lim\limits_{(x_1 \dots x_k)\to(+\infty,\dots,+\infty)} F(x_1,\dots,x_k) = 1$. Por cuanto en el límite se trata de obtener la probabilidad $P(\vec{X} \in R^k)$ que obviamente es la unidad.

De la misma forma que para las variables aleatorias, una función de distribución de probabilidad podía utilizarse para inducir una probabilidad sobre los borelianos de R, y tener así la equivalencia entre ambos conceptos (medida de probabilidad inducida y función de distribución de probabilidad), también aquí podemos actuar de forma semejante pero, eso sí, tomando algunas precauciones impuestas por los cambios que se han producido en la situación. En otras palabras, la generalización no puede hacerse directamente sino que hay que generalizar adecuadamente el concepto de función de distribución de probabilidad. Veámoslo:

Definición: Sea $F(\vec{x})$, $\vec{x}=(x_1,\dots,x_k)$ una función definida sobre R^k. Diremos que $F(\vec{x})$ es una _función de distribución de probabilidad_ si verifica las siguientes condiciones:

C.1) $\Delta_{b-a}\{F(a)\} \geq 0$, si $a \leq b$

C.2) F es creciente y continua por la derecha en cada una de las componentes de $\vec{x}$.

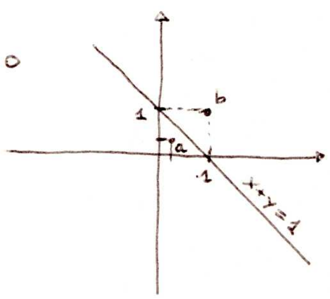
C.3) $\lim\limits_{x_i\to-\infty} F(x_1 \dots x_k) = 0$ para cada x_i.

C.4) $\lim\limits_{(x_1\dots x_k)\to(+\infty,\dots,+\infty)} F(x_1 \dots x_k) = 1$

Obsérvese que esta definición coincide con la dada para una función de distribución de probabilidad. Sin embargo ahora hemos debido especificar la condición C.1 por cuanto la monotonía y la continuidad no bastan para que $F(\vec{x})$ pueda utilizarse en la definición de una probabilidad sobre β^k de R^k. Consideremos el ejemplo que sigue para mejor comprender lo dicho.

Ejemplo 1.- Sea $F(x,y)$ definida como sigue

$$F(x,y)=\begin{cases}0, & \text{si } x<0, \text{ o } x+y<1, \text{ o } y<0\\ 1, & \text{en el resto}\end{cases}$$



Esta función satisface las condiciones C2, C3 y C4 de la anterior definición. Sean $a=(\frac13,\frac13)$, $b=(1,1)$, entonces

$$\Delta_{b-a}\{F(a)\} = F(1,1)+F(\tfrac13,\tfrac13)-F(1,\tfrac13)-F(\tfrac13,1) =$$
$$= 1+0-1-1=-1<0$$

por tanto no podemos utilizar $F(\vec{x})$ para la definición de una medida de probabilidad sobre β^k de R^k sobre la base de hacerlo mediante

$$P'(]a,b]) = \Delta_{b-a}\{F(a)\},$$

por cuanto nos encontraríamos con que $P'(]a,b]) = -1 < 0$, para $a=(\frac13,\frac13)$, $b=(1,1)$.

A partir de una función de distribución de probabilidad, como la definida, podemos definir una medida de probabilidad sobre β^k de R^k, haciéndolo en primera instancia sobre la familia de los intervalos de la forma $\{\vec{x};\ a<\vec{x}\leq b\} =]a,b]$, mediante

$$P(]a,b]) = \Delta_{b-a}\{F(a)\}.$$

A continuación, mediante el teorema de extensión, podemos hacer que la medida P trabaje sobre β^k. En definitiva esto viene a poner de manifiesto la equivalencia entre ambos conceptos a la hora de conocer las características de la distribución de probabilidad del vector aleatorio $\vec{X}$.

Existen, al igual que ocurría con las variables aleatorias, diversos tipos de vectores aleatorios. Nosotros tan sólo nos ocuparemos de los dos más interesantes y que son los equivalentes a las variables aleatorias estudiadas anteriormente.

Tipos de vectores aleatorios. Como hacíamos en el caso unidimensional podemos ahora introducir una nueva función, a saber

$$f(\vec{x}) = f(x_1,\dots,x_k) = P(\vec{X}=\vec{x}) = P(X_1=x_1,\dots,X_k=x_k).$$

Sea $A=\{\vec{x}\in R^k;\ f(\vec{x})>0\}$, si A es tal que

$$\sum_{\vec{x}\in A} f(\vec{x}) = 1$$

decimos entonces que $\vec{X}$ es un _vector aleatorio discreto_. Obsérvese que tal y como ocurría en el caso unidimensional A es a lo sumo numerable. A la función f se la denomina _función de cuantía_ y guarda con F la siguiente relación:

$$F(\vec{x}) = \sum_{\vec{t}\in A\cap]-\infty,\vec{x}]} f(\vec{t})$$

lo que supone que la $F(\vec{x})$ es el equivalente k-dimensional a una función en escalera en R.

Un razonamiento similar al utilizado en el caso unidimensional, que no reiteraremos aquí, nos lleva a introducir el concepto de _vector aleatorio continuo_, como aquel para el que existe una función f, tal que, siendo no negativa,

$$F(\vec{x}) = \int_{-\infty}^{\vec{x}} f(\vec{t})\,d\vec{t} = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k} f(t_1,\dots,t_k)\,dt_1\dots dt_k.$$

a la función f se la designa como _función de densidad de probabilidad_. Todas las consideraciones de tipo geométrico que en su momento hicimos para las variables aleatorias son aplicables ahora, sustituyendo hipervolumen k-dimensional donde decíamos área, etc...

La función de densidad de probabilidad tiene, como consecuencia de la definición, las propiedades siguientes:

a) Si f es continua en $(x_1,x_2,\ldots,x_k)$, entonces

$$f(x_1,\ldots,x_k) = \frac{\partial^k F(x_1,\ldots,x_k)}{\partial x_1 \ldots \partial x_k}$$

b)

$$\lim_{(x_1\ldots x_k)\to(+\infty,\ldots,+\infty)} F(x_1\ldots x_k) = 1 = \lim_{(x_1\ldots x_k)\to(+\infty,\ldots,+\infty)} \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k} f(t_1\ldots t_k)\,dt_1\ldots dt_k =$$

$$= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1\ldots x_k)\,dx_1\ldots dx_k .$$

De acuerdo con todo lo expuesto, y dado que la distribución de probabilidad de un vector aleatorio $\vec{X}$ podemos conocerla a través de su función de distribución, en última instancia también podemos conocer, y por tanto caracterizar, dicha distribución a partir de la función de densidad de probabilidad.

Ejemplo 2.- Sea X una variable aleatoria cuyo valor designa la cara que nos muestra un dado al ser lanzado. Sea Y una variable aleatoria que toma los valores 0 ó 1, según que una moneda, previamente lanzada, nos muestre su cruz o su cara. Tanto el dado como la moneda son perfectos y los lanzamientos independientes. Definimos un vector aleatorio mediante (X,Y). Dicho vector será del tipo discreto con

$$A=\{(1,0),(2,0),\ldots,(6,0),(1,1),\ldots,(6,1)\}$$

y $f(x_1,x_2)=\frac{1}{12}$ para $(x_1,x_2)\in A$.

La función de distribución viene dada por

$$F(x,y)=\begin{cases} 0 & x<1,\ -\infty<y<+\infty;\ -\infty<x<+\infty,\ y<0 \\ 1/12 & 1\le x<2,\ 0\le y<1 \\ 1/6 & 2\le x<3,\ 0\le y<1;\ 1\le x<2,\ 1\le y \\ 1/4 & 3\le x<4,\ 0\le y<1 \\ 1/3 & 4\le x<5,\ 0\le y<1;\ 2\le x<3,\ 1\le y \\ 5/12 & 5\le x<6,\ 0\le y<1 \end{cases}$$

$$F(x,y)=\begin{cases} 1/2 & 6\le x,\ 0\le y<1;\ 3\le x<4,\ 1\le y \\ 2/3 & 4\le x<5,\ 1\le y \\ 5/6 & 5\le x<6,\ 1\le y \\ 1 & 6\le x,\ 1\le y \end{cases}$$

Ejemplo 3.- Sea (X,Y) un vector aleatorio del tipo continuo con función de densidad de probabilidad dada por

$$f(x,y)=\begin{cases} e^{-(x+y)} & 0<x<+\infty,\ 0<y<+\infty \\ 0, & \text{en el resto.} \end{cases}$$

Efectivamente $f(x,y)$ es una f.d.p. por cuanto

a) $f(x,y)\ge 0 \quad \forall (x,y)$

b)

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_0^{+\infty}\int_0^{+\infty} e^{-(x+y)}\,dx\,dy = \left[(1-e^{-x})\right]_0^{+\infty}\int_0^{+\infty} e^{-y}\,dy =$$

$$= \left[1-e^{-y}\right]_0^{+\infty} = 1$$

Así pues la función de distribución del vector aleatorio (X,Y) será

$$F(x,y)=\int_0^{x}\int_0^{y} e^{-(t+u)}\,dt\,du = (1-e^{-x})(1-e^{-y}) \qquad 0<x<+\infty,\ 0<y<+\infty$$

$F(x,y)=0$ en el resto.

Funciones de distribución y de densidad marginales

Sea $\vec{X}$ un vector aleatorio y sea $F(\vec{x})$ su función de distribución. Consideremos los dos subvectores $\vec{X}_1$, $\vec{X}_2$ de dimensiones m y $k-m$ respectivamente, $m\ge 1$, de manera que $\vec{X}=(\vec{X}_1,\vec{X}_2)$ ya sabemos que como tales vectores también serán aleatorios. Consideremos la función de distribución en un pto. genérico $\vec{x}=(\vec{x}_1,\vec{x}_2)$ y supongamos que $\vec{x}_2\to+\infty$ en todas sus componentes, tendremos

$$F(\vec{x}_1,+\infty,\ldots,+\infty) = \lim_{\vec{x}_2\to(+\infty,\ldots,+\infty)} F(\vec{x}_1,\vec{x}_2) = \lim_{\vec{x}_2\to(+\infty,\ldots,+\infty)} P(\vec{X}_1\le\vec{x}_1,\vec{X}_2\le\vec{x}_2) = P(\vec{X}_1\le\vec{x}_1) =$$

$$= F_{\vec{X}_1}(\vec{x}_1)$$

es decir, se trata de la función de distribución del subvector $\vec{X}_1$. Actuando de esta manera podemos obtener la función de distribución de cualquier subvector de dimensión $m<k$. En particular para $m=1$, tendríamos las funciones de distribución de las variables aleatorias que componen el vector $\vec{X}$, a saber

$$F_i(x_i) = F(+\infty,\ldots,x_i,\ldots,+\infty).$$

Obsérvese que para cada dimensión, m, son exactamente $\binom{k}{m}$ el número de subvectores de dimensión m que pueden obtenerse. A todas estas funciones de distribución, así obtenidas, se les conoce con el nombre de funciones de distribución marginales (del subvector correspondiente $\vec{X}_1$). Por contraposición con esta nomenclatura a $F(\vec{x}) = F(x_1 \ldots x_k)$ se le conoce con el nombre de función de distribución conjunta.

Si $\vec{X}$ es un vector aleatorio discreto, recordemos que, si A es tal que $P(\vec{X}\in A)=1$, tenemos

$$F(\vec{x}) = \sum_{\vec{t}\in A\cap]-\infty,\vec{x}]} f(\vec{t}), \quad \text{con } f(\vec{t}) = P(\vec{X}=\vec{t})$$

en términos de los subvectores $\vec{X}_1$ y $\vec{X}_2$ la anterior expresión podría escribirse

$$F(\vec{x}_1,\vec{x}_2) = \sum_{(\vec{t}_1,\vec{t}_2)\in A\cap]-\infty,\vec{x}]} f(\vec{t}_1,\vec{t}_2) = \sum_{\vec{t}_1\in A_1\cap]-\infty,\vec{x}_1]}\left[\sum_{\vec{t}_2\in A_2\cap]-\infty,\vec{x}_2]} f(\vec{t}_1,\vec{t}_2)\right]$$

donde A_i es la proyección de A en el subespacio correspondiente, al igual que $]-\infty,\vec{x}_i]$, $i=1,2$. En esta situación

$$F_{\vec{X}_1}(\vec{x}_1) = \lim_{\vec{x}_2\to+\vec{\infty}} F(\vec{x}_1,\vec{x}_2) = \lim_{\vec{t}_2\to+\vec{\infty}}\left\{\sum_{\vec t_1\in A_1\cap]-\infty,\vec x_1]}\left[\sum_{\vec t_2\in A_2\cap]-\infty,\vec x_2]} f(\vec t_1,\vec t_2)\right]\right\} =$$

$$= \sum_{\vec t_1\in A_1\cap]-\infty,\vec x_1]}\left[\sum_{\vec t_2\in A_2} f(\vec t_1,\vec t_2)\right] = \sum_{\vec t_1\in A_1\cap]-\infty,\vec x_1]} f_{\vec X_1}(\vec t_1)$$

donde $f_{\vec X_1}(\vec x_1) = \sum_{\vec t_2\in A_2} f(\vec x_1,\vec t_2)$ es la función de cuantía del vector aleatorio $\vec X_1$, como fácilmente se desprende de las últimas igualdades escritas. Al igual que a su correspondiente función de distribución también a ésta se le conoce como la ~~marginal con~~ función de cuantía marginal, mientras que a la $f(\vec x)=f(x_1\ldots x_k)$ se la denomina función de cuantía conjunta.

Supongamos que el vector $\vec X$ es continuo, es decir, existe f tal que

$$F(\vec x) = F(x_1\ldots x_k) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k} f(t_1\ldots t_k)\,dt_1\ldots dt_k$$

o bien, adoptando una nomenclatura coherente con los subvectores $\vec X_1$ y $\vec X_2$,

$$F(\vec x_1,\vec x_2) = \int_{\vec t_1\le\vec x_1}\int_{\vec t_2\le \vec x_2} f(\vec t_1,\vec t_2)\,d\vec t_2\,d\vec t_1 .$$

Para la obtención de la función de distribución marginal de $\vec X_1$, tendremos

$$F_{\vec X_1}(\vec x_1) = \lim_{\vec x_2\to+\vec\infty} F(\vec x_1,\vec x_2) = \int_{\vec t_1\le \vec x_1}\left[\int_{\vec t_2\in R^{k-m}} f(\vec t_1,\vec t_2)\,d\vec t_2\right]d\vec t_1$$

De acuerdo con esto lo que figura dentro del corchete debe de ser la función de densidad marginal del subvector aleatorio $\vec X_1$, es decir

$$f_{\vec X_1}(\vec x_1) = \int_{\vec x_2\in R^{k-m}} f(\vec x_1,\vec x_2)\,d\vec x_2 = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1\ldots x_m, x_{m+1}\ldots x_k)\,dx_{m+1}\cdots dx_k .$$

Veamos un par de ejemplos.

Ejemplo 4. Lanzamos tres veces una moneda. Sea X la variable que representa el número de caras en los tres lanzamientos e Y la diferencia, en valor absoluto, entre el número de caras y el número de cruces. Obtener las funciones de cuantía conjunta y marginales y las de distribución.

Para mejor trabajar dispondremos los valores en la siguiente tabla:

$Y \backslash X$	0	1	2	3	$P(Y=y)$
1	0	3/8	3/8	0	6/8
3	1/8	0	0	1/8	2/8
$P(X=x)$	1/8	3/8	3/8	1/8	1

La función de cuantía conjunta son los valores que aparecen en el centro de la tabla. Las marginales las que aparecen en los márgenes de la misma.

Las funciones de distribución se obtendrán sumando adecuadamente. Así por ejemplo, para la conjunta

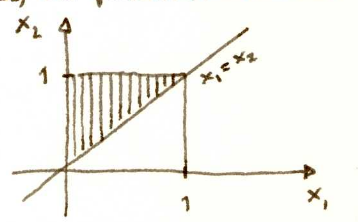
$$F(2.5,1.5) = P(X\le 2.5, Y\le 1.5) = \sum_{\substack{x\le 2.5\\ y\le 1.5}} f(x,y) = f(0,1)+f(1,1)+f(2,1) = 6/8$$

Para la marginal de la X

$$F_X(2.5) = P(X\le 2.5) = P(X=0)+P(X=1)+P(X=2) = 7/8 .$$

Ejemplo 5. Supongamos un vector aleatorio bidimensional (X_1,X_2) con función de densidad de probabilidad conjunta

$$f(x_1,x_2)=\begin{cases}2, & 0<x_1<x_2<1\\ 0, & \text{en el resto}\end{cases}$$



Las funciones de densidad marginales vienen dadas por

$$f_1(x_1) = \int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_2 = \int_{x_1}^{1} 2\,dx_2 = 2(1-x_1) \qquad 0<x_1<1$$

$$f_1(x_1)=0 \qquad \text{en el resto}$$

$$f_2(x_2)=\int_0^{x_2} 2\,dx_1 = 2x_2 \quad , \quad 0<x_2<1$$

$$f_2(x_2)=0 \quad \text{en el resto.}$$

La función de distribución conjunta viene dada por

$$F(x_1,x_2)=0 \quad \text{si} \quad x_1\leq 0 \ \text{ó} \ x_2\leq 0$$

$$F(x_1,x_2)=\int_0^{x_1}\int_{t_1}^{x_2} f(t_1,t_2)\,dt_2\,dt_1=\int_0^{x_1}\int_{t_1}^{x_2} 2\,dt_2\,dt_1 = 2\int_0^{x_1}(x_2-t_1)\,dt_1 = 2\left(x_2x_1-\frac{x_1^2}{2}\right)=$$

$$= x_1(2x_2-x_1) \qquad 0<x_1<x_2<1$$

$$F(x_1,x_2) = 2x_1-x_1^2 \quad , \quad 0<x_1<1,\ x_2\geq 1$$

$$F(x_1,x_2) = x_2^2 \quad , \quad 0<x_2<1,\ x_1\geq x_2$$

$$F(x_1,x_2) = 1 \quad , \quad x_1\geq 1,\ x_2\geq 1.$$

Las marginales son, para X_1

$$F_{X_1}(x_1)=\int_0^{x_1}(2-2t_1)\,dt_1 = 2x_1-x_1^2 \quad , \quad 0<x_1<1$$

$$F_{X_1}(x_1)=1, \quad x_1\geq 1$$

$$F_{X_1}(x_1)=0 \quad x_1\leq 0$$

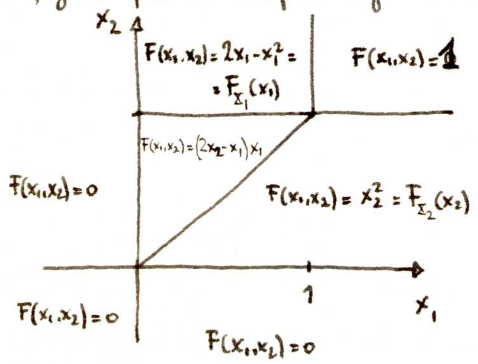
para X_2

$$F_{X_2}(x_2)=0, \quad x_2\leq 0$$

$$F_{X_2}(x_2)=\int_0^{x_2} 2t_2\,dt_2 = x_2^2 \quad , \quad 0<x_2<1$$

$$F_{X_2}(x_2)=1, \quad x_2\geq 1$$

La figura pone de manifiesto conjuntamente, todos estos resultados



Funciones de densidad y de distribución condicionadas.

Caso discreto. Sea $\vec{X}$ un vector aleatorio discreto y considerémoslo nuevamente fraccionado en dos subvectores $(\vec{X}_1,\vec{X}_2)=\vec{X}$. Sea además $\vec{X}_1=\vec{x}_1$ de manera que $f_{\vec{X}_1}(\vec{x}_1)>0$, definamos una nueva función, del subvector $\vec{X}_2$, mediante

$$f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)=\frac{f(\vec{x}_1,\vec{x}_2)}{f_{\vec{X}_1}(\vec{x}_1)}$$

Así definida esta función tiene las siguientes propiedades:

a) $f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)\geq 0 \quad , \quad \forall \vec{x}_2$

b) $$\sum_{\vec{x}_2} f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)=\sum_{\vec{x}_2}\frac{f(\vec{x}_1,\vec{x}_2)}{f_{\vec{X}_1}(\vec{x}_1)}=\frac{1}{f_{\vec{X}_1}(\vec{x}_1)}\left[\sum_{\vec{x}_2} f(\vec{x}_1,\vec{x}_2)\right]=$$

$$=\frac{1}{f_{\vec{X}_1}(\vec{x}_1)}\cdot f_{\vec{X}_1}(\vec{x}_1)=1$$

pero estas son las condiciones que definen a las funciones de densidad de probabilidad, es decir $f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)$ es una f.d.p., pero ¿qué significado tiene? Veámoslo.

Recordemos que $f(\vec{x}_1,\vec{x}_2)=P(\vec{X}_1=\vec{x}_1,\vec{X}_2=\vec{x}_2)$ y $f_{\vec{X}_1}(\vec{x}_1)=P(\vec{X}_1=\vec{x}_1)$, así pues

$$f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)=\frac{P(\vec{X}_1=\vec{x}_1,\vec{X}_2=\vec{x}_2)}{P(\vec{X}_1=\vec{x}_1)}=P(\text{[redacted]},\vec{X}_2=\vec{x}_2/\vec{X}_1=\vec{x}_1)$$

La función representa pues la probabilidad condicional de que el subvector $\vec{X}_2$ tome el valor $\vec{x}_2$ si el subvector $\vec{X}_1$ tomó el valor $\vec{x}_1$. Obsérvese además que

$$P(\vec{X}_2\in B/\vec{X}_1=\vec{x}_1)=\sum_{\vec{x}_2\in B}P(\vec{X}_2=\vec{x}_2/\vec{X}_1=\vec{x}_1)=\sum_{\vec{x}_2\in B} f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1).$$

En definitiva, $f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)$ es la función de cuantía condicionada de $\vec{X}_2$, dado que $\vec{X}_1=\vec{x}_1$.

Análogamente se define $f_{\vec{X}_1/\vec{X}_2}(\vec{x}_1/\vec{x}_2)$ siempre que $\vec{x}_2$ sea tal que $f_{\vec{X}_2}(\vec{x}_2)>0$.

Ejemplo 6. Recordemos el vector aleatorio bidimensional del ejemplo 4. Algunas de las condicionadas que allí pueden definirse serían, por ejemplo

$$f_{X/Y=1}(x/1) = 0 \quad , \text{ para } x=0,\ x=3$$

$$f_{X/Y=1}(x/1) = \tfrac{1}{2} \quad , \text{ para } x=1,\ x=2$$

$$f_{X/Y=3}(x/3) = \tfrac{1}{2} \quad , \text{ para } x=0,\ x=3$$

$$f_{X/Y=3}(x/3) = 0 \quad , \text{ para } x=1,\ x=2$$

$$f_{Y/X=0}(y/0) = 0 \quad , \text{ para } y=1$$

$$f_{Y/X=0}(y/0) = 1 \quad , \text{ para } y=3$$

El vector aleatorio condicionado $\vec{X}_2/\vec{X}_1=\vec{x}_1$ tiene como función de distribución condicionada a $F_{\vec{X}_2/\vec{x}_1}(\vec{x}_2/\vec{x}_1)$, cuyo valor, en función de la función de cuantía condicionada es

$$F_{\vec{X}_2/\vec{x}_1}(\vec{x}_2/\vec{x}_1) = \sum_{\vec{t}_2 \le \vec{x}_2} f_{\vec{X}_2/\vec{x}_1}(\vec{t}_2/\vec{x}_1).$$

Así por ejemplo, en el caso anterior tendríamos

$$F_{X/Y=1}(2/1) = \sum_{t\le 2} f_{X/Y=1}(t/1) = f_{X/Y=1}(0/1) + f_{X/Y=1}(1/1) + f_{X/Y=1}(2/1) = 0 + \tfrac{1}{2} + \tfrac{1}{2} = 1.$$

Caso continuo.- Supongamos ahora que $\vec{X}$ es un vector aleatorio del tipo continuo y para facilitar el desarrollo que vendrá a continuación vamos a situarnos en el caso bidimensional, es decir, $\vec{X}=(X,Y)$. Recordemos que en el caso continuo $P(X=x)=0$ y $P(Y=y)=0$ y por tanto ahora, probabilidades del tipo $P(X\le x/Y=y)$ no están definidas. Sea entonces $\varepsilon>0$ y supongamos que $P(y-\varepsilon<Y\le y+\varepsilon)>0$, definiremos entonces la siguiente probabilidad condicional,

$$P(X\le x/y-\varepsilon<Y\le y+\varepsilon) = \frac{P(X\le x,\ y-\varepsilon<Y\le y+\varepsilon)}{P(y-\varepsilon<Y\le y+\varepsilon)}.$$

Se trata, para todo intervalo fijo $]y-\varepsilon, y+\varepsilon]$, la probabilidad condicional de que $X\le x$, dado que $Y\in\,]y-\varepsilon,y+\varepsilon]$, es decir, de la función de distribución condicional de X, dado que $Y\in\,]y-\varepsilon,y+\varepsilon]$.

Consideremos ahora el siguiente límite

$$\lim_{\varepsilon\to 0^+} P(X\le x/Y\in\,]y-\varepsilon,y+\varepsilon]).$$

Cuando dicho límite existe lo designaremos mediante $F_{X/Y}(x/y)$ y lo definiremos como la función de distribución condicional de X, dado $Y=y$. Bajo las condiciones ya establecidas de continuidad absoluta de la función de distribución considerada, sabemos que existe una función no negativa que verifica

$$F_{X/Y}(x/y) = \int_{-\infty}^{x} g(t)\,dt \quad , \forall x\in\mathbb{R}.$$

a esta función g la designaremos como función de densidad de probabilidad condicional de X, dado $Y=y$, y la denotaremos mediante $f_{X/Y}(x/y)$. Es más que evidente que la función g verifica las condiciones que definen a una f.d.p.

Llegados a este punto cabe preguntarnos qué relación existe entre la f.d.p. condicionada y las f.d.p. conjuntas y marginales del vector (X,Y). Veámoslo mediante el siguiente

TEOREMA.- Sea f la f.d.p. conjunta del vector aleatorio continuo (X,Y), y sea f_2 la marginal correspondiente de Y. Para cada punto de continuidad de f en el que además $f_2(y)>0$ y es continua, se verifica

$$f_{X/Y}(x/y) = \frac{f(x,y)}{f_2(y)}.$$

Demostración.- Recordemos que

$$F_{X/Y}(x/y) = \lim_{\varepsilon\to 0^+} \frac{P(X\le x,\ y-\varepsilon<Y\le y+\varepsilon)}{P(y-\varepsilon<Y\le y+\varepsilon)} = \lim_{\varepsilon\to 0^+} \frac{\int_{-\infty}^{x}\left\{\int_{y-\varepsilon}^{y+\varepsilon} f(u,v)\,du\,dv\right\}}{\int_{y-\varepsilon}^{y+\varepsilon} f_2(v)\,dv}$$

dividiendo numerador y denominador por 2ε, pasando al límite y teniendo en cuenta la relación que liga a las f.d.p. con las funciones de distribución en los puntos de continuidad de aquellas, tendremos:

$$F_{X/Y}(x/y) = \frac{\int_{-\infty}^{x} f(u,y)\,du}{f_2(y)} = \int_{-\infty}^{x}\left\{\frac{f(u,y)}{f_2(y)}\right\}du$$

si derivamos ahora, respecto de x, la función de distribución condicional, tendremos,

$$f_{X/Y}(x/y) = \frac{f(x,y)}{f_2(y)}$$

que es el resultado buscado.

Observación.- De la igualdad obtenida en el teorema

$$F_{X/Y}(x/y) = \frac{\int_{-\infty}^{x} f(u,y)\,du}{f_2(y)}$$

tenemos

$$f_2(y)\cdot F_{X/Y}(x/y) = \int_{-\infty}^{x} f(u,y)\,du$$

Si integramos para y en toda la recta real, tendríamos

$$\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{x} f(u,y)\,du\right]dy = \int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(u,y)\,dy\right]du = \int_{-\infty}^{x} f_1(u)\,du = F_1(x)$$

que es la función de distribución marginal de X, y en la otra parte del igual quedaría

$$F_1(x) = \int_{-\infty}^{+\infty} f_2(y)\cdot F_{X/Y}(x/y)\,dy.$$

Estas definiciones pueden trasladarse de forma inmediata al caso de un espacio m-dimensional. En efecto, para $\vec{X} = (\vec{X}_1, \vec{X}_2)$, tendríamos

$$F_{\vec{X}_1/\vec{X}_2}(\vec{x}_1/\vec{x}_2) = \frac{\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k} f(x_1,\dots,x_m)\,dx_1\dots dx_k}{f_{\vec{X}_2}(x_{k+1},\dots,x_m)}$$

y para la función de densidad de probabilidad condicional

$$f_{\vec{X}_1/\vec{X}_2}(\vec{x}_1/\vec{x}_2) = \frac{f(\vec{x})}{f_{\vec{X}_2}(\vec{x}_2)}, \quad \text{con } \vec{x} = (\vec{x}_1, \vec{x}_2).$$

Ejemplo 7.- Utilizando el vector bidimensional del ejemplo 5 (pag. 8), tendríamos

$$f_{X_1/X_2}(x_1/x_2) = \frac{f(x_1,x_2)}{f_2(x_2)} = \frac{2}{2x_2} = \frac{1}{x_2}, \quad 0<x_1<x_2$$

es decir, se trata de una distribución uniforme en el intervalo $]0,x_2[$.

Análogamente

$$f_{X_2/X_1}(x_2/x_1) = \frac{f(x_1,x_2)}{f_1(x_1)} = \frac{2}{2(1-x_1)} = \frac{1}{1-x_1}, \quad x_1<x_2<1$$

que es, nuevamente, una distribución uniforme en el intervalo $]x_1,1[$. Para las funciones de distribución correspondientes

$$F_{X_1/X_2=x_2}(x_1/x_2) = \int_0^{x_1} f_{X_1/X_2}(t/x_2)\,dt = \int_0^{x_1}\frac{1}{x_2}\,dt = \frac{x_1}{x_2}, \quad 0<x_1<x_2$$

para $x_2 = \frac{2}{3}$, $x_1 = \frac{1}{3}$, tendríamos

$$F_{X_1/X_2=\frac{2}{3}}(1/3\,|\,2/3) = P\left(X_1 \le \tfrac{1}{3} \,/\, X_2 = \tfrac{2}{3}\right) = \frac{1/3}{2/3} = \frac{1}{2}.$$

Por otra parte

$$F_1(x_1) = \int_{-\infty}^{+\infty} f_2(x_2)\cdot F_{X_1/X_2=x_2}(x_1/x_2)\,dx_2 = \int_{x_1}^{1} 2x_2\left[\frac{x_1}{x_2}\right]dx_2 = x_1\int_{x_1}^{1} 2\,dx_2 = x_1\left[2x_2\right]_{x_1}^{1} =$$

$$= 2x_1 - 2x_1^2, \quad 0<x_1<1.$$

$$F_1(x_1) = 0, \quad x_1 \le 0$$

$$F_1(x_1) = 1, \quad x_1 \ge 1.$$

como ya sabíamos.

Truncamiento de una variable aleatoria

Hemos visto ~~que~~ que si X es una variable aleatoria continua, mediante el cambio $Y=g(X)$ con $g(x)=x$, para $|x|<b$, $g(x)=-b$, $x\le -b$ y $g(x)=b$, $x\ge b$, podíamos llevar a cabo un truncamiento de la variable X, pero existía el inconveniente de obtener una nueva variable, Y, que no era ni discreta, ni continua, por cuanto su función de distribución no era continua, ni tampoco escalera. Haciendo uso de las probabilidades condicionadas podemos llevar a cabo un truncamiento de la variable que resuelve este inconveniente. Veamos como.

Definición.- Si X es una variable aleatoria [redacted] sobre el espacio de probabilidad $(S, \mathcal{A}, P)$ y si T es un boreliano de la recta real, tal que $0<P(X\in T)<1$, entonces a la distribución condicional $P(X\le x/X\in T)$, definida para cualquier x, se le llama distribución truncada de X.

Obsérvese que la definición es válida para cualquier tipo de variable, así, para el caso de una variable discreta, las cosas quedan de la siguiente manera:

$$P(X=x_i/X\in T) = \frac{P(X=x_i, X\in T)}{P(X\in T)} = \begin{cases} \dfrac{f(x_i)}{\sum_{x_j\in T} f(x_j)}, & \text{si } x_i\in T \\ 0, & \text{en el resto} \end{cases}$$

El caso más interesante, aquel en que X es continua, da lugar a

$$P(X\le x/X\in T)=\frac{P(X\le x, X\in T)}{P(X\in T)}=\frac{\int_{]-\infty,x]\cap T} f(t).dt}{\int_T f(t).dt}$$

La función de densidad de probabilidad para la variable truncada es

$$h(x)=\begin{cases}\dfrac{f(x)}{\int_T f(t).dt} & , x\in T\\ 0 & , x\in T^c .\end{cases}$$

El efecto del truncamiento es acumular sobre T toda la masa de probabilidad, por cuanto

$$P(X\in T/X\in T)=1.$$

Este es un método muy utilizado en teoría de la probabilidad y muy útil en particular cuando se estudian teoremas límite. Otra interesante aplicación del método de truncamiento de una variable sobreviene cuando la variable original no tiene esperanza finita, pero de ello nos ocuparemos en el próximo capítulo.

Ejemplo 8.- Sea X una v.a. $N(0,1)$ y sea $T=]-\infty,0]$. Entonces $P(X\in T)=1/2$ debido a la simetría de X. Para la variable truncada tendremos pues

$$h(x)=\begin{cases}2f(x)=\dfrac{\sqrt{2}}{\sqrt{\pi}}.e^{-\frac{1}{2}x^2} & , x\le 0\\ 0 & , \text{en el resto.}\end{cases}$$

Algunos vectores aleatorios usuales.

$\boxed{\text{Discretos}}$

1. Distribución multinomial.- Se trata de una generalización de la variable aleatoria binomial que en su momento estudiamos. Surge en situaciones como la siguiente:

Supongamos que llevamos a cabo n pruebas, independientes unas de otras, de tal manera que en cada una de ellas podemos tener como resultado uno de los r sucesos $E_1, E_2, \dots, E_r$, cada uno de los cuales teniendo una probabilidad p_i, $i=1,\dots,r$ de ocurrir, verificándose además $p_1+p_2+\dots+p_r=1$. En estas condiciones la probabilidad de obtener, al final de las n pruebas, k_1 veces E_1, k_2 veces E_2, ..., k_r veces E_r, con $k_1+k_2+\dots+k_r=n$ vendrá dada por:

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-(k_1+k_2+\dots+k_{r-1})}{k_r}\, p_1^{k_1}\cdot p_2^{k_2}\cdots p_r^{k_r}$$

que también puede escribirse

$$\frac{n!}{k_1!\,k_2!\cdots k_r!}\, p_1^{k_1}\cdots p_r^{k_r}.$$

Definamos ahora un vector aleatorio $\vec{X}=(X_1,\dots,X_r)$ donde X_i designa el número de veces que ocurre el [redacted] suceso E_i. Para $\vec{x}\in A$, siendo $A=\{\vec{x}\in R^r ;\ x_i\ge 0, i=1\dots r \wedge \sum_{i=1}^r x_i=n\}$, tendremos que la función de cuantía vale

$$f(\vec{x})=P(X_1=x_1,\dots,X_r=x_r)=\frac{n!}{x_1!\cdots x_r!}\,p_1^{x_1}\cdots p_r^{x_r}$$

mientras que si $\vec{x}\in A^c$, entonces

$$f(\vec{x})=0$$

La función así definida es efectivamente una función de cuantía, por cuanto es no negativa en R^r y además, al sumarla para todos los elementos de A, y teniendo en cuenta el significado de los coeficientes multinomiales $\left(\frac{n!}{x_1!\dots x_r!}\right)$, tendremos

$$\sum_{\vec{x}\in A} f(\vec{x})=\sum_{\vec{x}\in A}\frac{n!}{x_1!\dots x_r!}\,p_1^{x_1}\dots p_r^{x_r}=(p_1+p_2+\dots+p_r)^n=1^n=1.$$

Obtención de algunas marginales.- Si queremos obtener la marginal de X_1, procederemos así

$$f_1(x_1)=\sum_{\vec{X}_2\in A_2} f(\vec{x}) \quad \text{con } \vec{X}_2=(x_2,\dots,x_r) \text{ y } A_2=\{\vec{x}_2 ;\ x_i\ge 0, i=2,\dots,r,\ \sum_{i=2}^r x_i=n-x_1\}$$

entonces

$$f_1(x_1) = \sum_{\vec{x}_2 \in A_2} f(x_1, \vec{x}_2) = \sum_{\vec{x}_2 \in A_2} \binom{n}{x_1}\binom{n-x_1}{x_2} \cdots \binom{x_r}{x_r} p_1^{x_1} \cdots p_r^{x_r} =$$

$$= \binom{n}{x_1} p_1^{x_1} \sum_{\vec{x}_2 \in A_2} \binom{n-x_1}{x_2}\binom{n-(x_1+x_2)}{x_3} \cdots \binom{x_r}{x_r} p_2^{x_2} \cdots p_r^{x_r} =$$

$$= \binom{n}{x_1} p_1^{x_1} (p_2 + \cdots + p_r)^{n-x_1} = \binom{n}{x_1} p_1^{x_1} (1-p_1)^{n-x_1}$$

se trata pues de una binomial con parámetros n, p_1. Obsérvese que razonando intuitivamente sobre el significado de la marginal se espera el resultado encontrado.

Para marginales que cualquier otra dimensión se obtienen nuevamente multinomiales de dimensión coincidente con la del vector marginal considerado.

Para las distribuciones condicionadas llegaríamos a un resultado análogo.

2. Distribución binomial negativa en el caso bivariante. Supongamos que E_1, E_2, E_3 constituyen una partición del espacio muestral, de manera que $P(E_1) = p_1$, $P(E_2) = p_2$, $P(E_3) = p_3 = 1 - p_1 - p_2$. Realizamos un experimento cuyo resultado es necesariamente E_1, E_2 ó E_3. Vamos a repetir este experimento, de manera independiente en cada ocasión, hasta que el suceso E_3 se haya realizado k veces, ello significa que E_1 habrá ocurrido x veces y E_2 habrá ocurrido y veces y todo ello antes de la k-ésima ocurrencia del suceso E_3. La probabilidad de que las cosas hayan ocurrido así, vendrá dada por

$$\binom{x+y+k-1}{x} \cdot \binom{y+k-1}{y} \cdot p_1^x \cdot p_2^y \cdot (1-p_1-p_2)^k = \frac{(x+y+k-1)!}{x!\, y!\, (k-1)!} p_1^x \cdot p_2^y \cdot (1-p_1-p_2)^k$$

Definamos ahora un vector aleatorio bidimensional (X, Y) donde X e Y designan el número de ocurrencias de los sucesos E_1 y E_2 antes de la k-ésima ocurrencia del suceso E_3. La función de cuantía de este vector aleatorio viene dada por

$$f(x,y) = P(X = x, Y = y) = \frac{(x+y+k-1)!}{x!\, y!\, (k-1)!} p_1^x \cdot p_2^y \cdot (1-p_1-p_2)^k, \quad (x,y) \in A$$

donde $A = \{(x,y) \in R^2;\ x = 0,1,2,\ldots\ \ y = 0,1,2,\ldots\}$, y

$$f(x,y) = 0, \quad (x,y) \in A^c.$$

Se trata de una función de cuantía por cuanto, $f(x,y) \geq 0$, $\forall (x,y) \in R^2$, además, teniendo en cuenta que

$$(1-p_1-p_2)^{-k} = \sum_{x,y \in A} \binom{x+y+k-1}{x}\binom{y+k-1}{y} p_1^x \cdot p_2^y$$

entonces

$$\sum_{(x,y) \in A} f(x,y) = (1-p_1-p_2)^k \cdot (1-p_1-p_2)^{-k} = 1.$$

Distribuciones marginales. La marginal de Y, por ejemplo, la estudiaremos como sigue:

$$f_2(y) = \sum_{x \in A_1} f(x,y) \qquad \text{donde } A_1 = \{x;\ x = 0,1,2,\ldots\}$$

entonces

$$f_2(y) = \sum_{x \geq 0} \binom{x+y+k-1}{x}\binom{y+k-1}{y} p_1^x p_2^y (1-p_1-p_2)^k =$$

$$= \binom{y+k-1}{y} p_2^y (1-p_1-p_2)^k \sum_{x \geq 0} \binom{x+y+k-1}{x} p_1^x$$

pero recordemos que $(1-x)^{-n} = \sum_{j \geq 0} \binom{n+j-1}{j} x^j$, por tanto

$$f_2(y) = \binom{y+k-1}{y} p_2^y (1-p_1-p_2)^k (1-p_1)^{-(y+k)} = \binom{y+k-1}{y} \left(\frac{p_2}{1-p_1}\right)^y \cdot \left(\frac{1-p_1-p_2}{1-p_1}\right)^k =$$

$$= \binom{y+k-1}{y} q_2^y (1-q_2)^k \qquad \text{con } q_2 = \frac{p_2}{1-p_1}, \text{ siendo } 0 < q_2 < 1, \ y = 0,1,\ldots$$

se trata pues de una binomial negativa con parámetros k, y $q_2 = \frac{p_2}{1-p_1}$. Razonando de forma análoga llegaríamos a

$$f_1(x) = \binom{x+k-1}{x} q_1^x (1-q_1)^k, \quad \text{con } q_1 = \frac{p_1}{1-p_2}, \quad x = 0,1,\ldots$$

Distribuciones condicionadas. Para $Y = y$, la variable aleatoria condicionada X/y, tiene como función de cuantía

$$f_{X/y}(x/y) = \frac{f(x,y)}{f_2(y)} = \frac{\binom{x+y+k-1}{x}\binom{y+k-1}{y} p_1^x p_2^y (1-p_1-p_2)^k}{\binom{y+k-1}{y}\left(\frac{p_2}{1-p_1}\right)^y \cdot \left(\frac{1-p_1-p_2}{1-p_1}\right)^k} =$$

$$= \binom{x+y+k-1}{x} p_1^x (1-p_1)^{y+k}, \quad x = 0,1,2,\ldots \quad y = 0,1,\ldots$$

que es una binomial negativa con parámetros $y+k$ y p_1.