Apuntes Estadística

Fernando Parra

3 de mayo de 2014

1. Variables aleatorias bidimensionales

1.1. Variables aleatoria bidimensionale

1.2. Distribución bidimensional discreta

Definición Sea (X, Y) una variable aleatoria bidimensional discreta con rango $R_{X\times Y}$. A cada posible resultado (x, y) de (X, Y), asociamos un número

$$p(x, y) = P[X = x, Y = y]$$

llamado la *función de probabilidad conjunta*, que cumple las siguientes condiciones

1. $p(x, y) \geq \forall(x, y)$
2. $\sum_{(x,y)} \sum_{\in R_{X\times Y}} p(x, y) = 1$

Los valores $[(x, y), p(x, y)]$ para todo $(x, y) \in R_{X\times Y}$, se llama la *distribución de probabilidad conjunta*. La probabilidad de un evento cualquiera A en $R_{X\times Y}(A \subset R_{X\times Y})$, está definido por

$$\boxed{P[(X, Y) \in A] = \sum_{(x,y)\in A} p(x, y)}$$

Como en el caso unidimensional, la distribución de probabilidad conjunta de (X, Y) es inducida por la probabilidad de los sucesos del espacio muestral original Σ. Si la variable aleatoria bidimensional es finita, la distribución de probabilidad cinjunta se representa en una tabla co dos entradas como se muestra en esta tabla.

$y \setminus x$	x_1	x_2	$\cdots$	x_n
y_1	$p(x_1, y_1)$	$p(x_2, y_1)$	$\cdots$	$p(x_n, y_1)$
y_2	$p(x_1, y_2)$	$p(x_2, y_2)$	$\cdots$	$p(x_n, y_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$	$\ldots$	$p(x_n, y_m)$

Definición La *función de distribución acumulada* de la variable aleatoria bidimensional está definida por

$$\boxed{F(x, y) = P[X \leq x, Y \leq y] = \sum_{u=-\infty}^{x} \sum_{v=-\infty}^{y} p(u, v)}$$

1.2.1. Distribuciones marginales

En algunos casos se puede estar interesado solo en las distribuciones de X o de Y.

Definición La función de probabilidad individual para X e Y se llaman *distribuciones de probabilidad marginal o funciones de probabilidad marginal.* La dsitribución marginal para X, está dado por

$$\boxed{p_X(x) = P[X = x] = \sum_{y \in R_Y|X=x} P[X = x, Y = x] = \sum_{y \in R_Y|X=x} p(x, y)}$$

donde $y \in R_Y|X = x$ es el rango de valores para Y, dado que $X = x$, note que este rango puede ser diferente para valores diferentes de x.
La *distribución marginal* para Y, esta dada por

$$\boxed{p_Y(y) = P[Y = y] = \sum_{x \in R_X|Y=y} p(x, y)}$$

El nombre marginal se debe a que los valores de estas distribuciones están dadas por la suma total en los márgenes de las filas y columnas de la representación tabular de $p(x, y)$.

1.2.2. Variables aleatorias independientes

El concepto de independencia de variables aleatorias es muy importante en estadística. Como en el caso de independencia de eventos intuitivamente diremos que las variables aleatorias X e Y son *independientes* si el resultado de una de las variables, digamos X no influye en el resultado de Y, y viceversa.

Definición Sean (X, Y) una variable aleatoria bidimensional discreta con función de probabilidad marginal $p_X(x)$ y $p_Y(y)$. Se dice que X e Y son independientes si, solo sí

$$p(x, y) = p_X(x)p_Y(y), \qquad \forall(x, y)$$

o que $P[X = x, Y = y] = P[X = x] = P[Y = y]$, $\forall(x, y)$. En otras palabras, X e Y son independientes si, $P[X = x]$ no afecta a la $P[Y = y]$.

1.2.3. Distribución de probabilidad condicional

Dado dos eventos A y B, la probabilidad condicional de que ocurra el evento A dado que ha ocurrido el evento B. Extenderemos este concepto a la distribución condicional de una variable aleatoria X, dado que la variable aleatoria Y haya tomado un valor determinado, por ejemplo y_0.
Sea (X, Y) un avariable aleatoria bidimensional con función de probabilidad $p(x, y)$ de rango $R_{X \times Y}$, y probabilidad marginales es $p_X(x)$, $p_Y(y)$. R_X y R_Y los rangos de las variables aleatorias X e Y respectivamente.
Sean los eventos $A = (X = x)$, $B = (Y = y)$, $(x, y) \in R_{X \times Y}$, entonces, por la definición de probabilidad condicional se tiene.

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

Pero, $A \cap B = (X = x, Y = y)$, por lo tanto $P[A \cap B] = P[X = x, Y = y] = p(x, y)$ y $P[B] = P[Y = y] = p_Y(y) > 0$, pues $y \in R_Y$.
Luego, la probabilidad condicional del evento A, dado B es

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{p[X = x, Y = y]}{P[Y = y]} = \frac{p(x, y)}{p_Y(y)}$$

Es decir

$$\boxed{P[X = x|Y = y] = \frac{p[X = x, Y = y]}{P[Y = y]} = \frac{p(x, y)}{p_Y(y)}}$$

La razón $\dfrac{p(x, y)}{p_Y(y)}$, se llama la función de *probabilidad condicional* de X, dado $Y = y$. La siguiente definición formaliza lo expuesto en el párrafo anterior.

Definición Si (X, Y) es una variable aleatoria bidimensional discreta, la *función de probabilidad condicional* de X dado $Y = y$, se denota por $p_{X|Y}(x|y)$ y se define por

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}\ , \qquad p_Y(y) > 0$$

Similarmente, la *función de probabilidad condicional* de Y dado $X = x$, se define por

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)}\ , \qquad p_X(x) > 0$$

Notamos que $p_{X|Y}(x|y) > 0$.

$$\sum_{x \in R_X|Y=y} p_{X|Y}(x|y) = \sum_{x \in R_X|Y=y} \frac{p(x,y)}{p_Y(y)} = \frac{\sum_{x \in R_X|Y=y} p(x,y)}{p_Y(y)} = \frac{p_Y(y)}{p_Y(y)} = 1$$

Es decir, $p_{X|Y}(x|y)$ cumple las condiciones de una función de probabilidad, entonces se puede calcular probabilidades tales como

$$P[a \leq X \leq b|Y = y] = \sum_{x \in R_X|a \leq x \leq b} p_{X|Y}(x|y)$$
$$P[X < a|Y = y] = \sum_{x \in R_X|a < x} p_{X|Y}(x|y)$$

Teorema Sea (X, Y) una variable aleatoria bidimensional, entonces

$$\begin{aligned} p_{X|Y}(x|y) = p_X(x)\ , &\qquad \forall(x,y) \\ p_{Y|X}(y|x) = p_Y(y)\ , &\qquad \forall(x,y) \end{aligned}$$

si solo si X e Y son variables aleatorias independientes.

1.2.4. Esperanza y varianza

En algunos problemas, se requiere calcular, el valor esperado de una función de dos o más variables aleatorias X e Y, por ejemplo: $X + Y$, XY, $X^2Y + X$. La técnia para obtener el valor esperado de una función de dos variables aleatorias es la misma que en el caso de una variables, como indica la definición siguiente.

Definición Sea (X, Y) una variable aleatoria bidimensional discreta, y $H(X, Y)$ una función de dos variables aleatorias X e Y. El *valor esperado* de $H(X, Y)$, se define por

$$E[H(X,Y)] = \sum_{R_{X \times Y}} \sum H(x,y)p(x,y)$$

Siempre que la suma sea absolutamente convergente. En el caso particular de que $H(X, Y) = X$, se tiene

$$\mu_X = E(X) = \sum_{R_{X \times Y}} \sum xp(x,y) = \sum_{x \in R_X} x \sum_{y \in R_Y} p(x,y) = \sum_{x \in R_X} xp_X(x)$$

Similarmente, si $H(X, Y) = Y$, $\mu_Y = E(Y) = \sum_{y \in R_Y} xp_Y(y)$. Es decir, la media de $X(Y)$, se determina, de la función de probabilidad marginal de $X(Y)$. El lector, siguiendo un proceso similar, puede encontrar que la varianza de X y de Y están dados por

$$\sigma_X^2 = E(X - \mu_X)^2 = \sum_{x \in R_X} (x - \mu_X)^2 p_X(x) = E(X^2) - (\mu_X)^2$$
$$\sigma_Y^2 = E(Y - \mu_Y)^2 = \sum_{y \in R_Y} (y - \mu_Y)^2 p_Y(y) = E(Y^2) - (\mu_Y)^2$$

observe que

$$H(X,Y) = (X - \mu_X)^2\ ; \quad H(X,Y) = (Y - \mu_Y)^2$$

Teorema Si (X, Y) es una variable aleatoria bidimensional, entonces

$$E[aH(x) \pm bG(Y)] = aE[H(X)] \pm bE[G(Y)]$$

una secuencia inmediata de este teorema, cuando

$$H(X) = X, \text{ y } G(Y) = Y, \text{ es: } E(X \pm Y) = E(X) \pm E(Y)$$

Teorema Si X e Y son dos variables aleatorias independientes, entonces

1. $E(XY) = E(X)E(Y)$
2. $Var(X + Y) = Var(X) + Var(Y)$

Una generalización de 1. de este teorema es la siguiente. Si X e Y son variables aleatorias independientes y $H(X)$, $G(Y)$ son funciones de X e Y respectivamente, entonces

$$E[H(X)G(Y)] = E[H(X)]E[G(Y)]$$

También 2. se puede generalizar a n variables aleatorias.

1.2.5. Esperanza condicional

1.2.6. Covarianza y coeficientes de correlación

Daremos la definición de unos números que miden como, están relacionados los valores posible de X con los valores posibles de Y; estos números dependen de la función de probabilidad conjunto de X e Y.

Definición Sean X e Y dos variables aleatorias con media μ_X y μ_Y, respectivamente. La *covarianza* de X e Y, denotado por $Cov(X, Y)$, σ_{XY}, se define por

$$Cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_{(x,y) \in R_{X \times Y}} (x - \mu_X)(y - \mu_y)p(x, y)$$

A diferencia de la variza que siempre es no negativa, la covarianza puede ser negativa, cero o positiva. Si los valores más pequeños de X son asociados con valores grandes Y, y viceversa, entonces la covarianza es negativa. Es decir, si $X - \mu_x$ e $Y - \mu_y$ tienen signos opuestos. En cambio la covarianza será positiva cuando los valores grandes de X, se asocian con valores grandes de Y, y valores pequeños de X son asociados con valores pequeños aunque pueden ser dependientes. Una fórmula alternativa para calcular covarianzas, da el siguiente teorema.

Teorema La *covarianza* de dos variables aleatorias X e Y, con media μ_X y μ_Y respectivamente, está dado por

$$Cov(X, Y) = E(XY) - \mu_X \mu_Y$$

Teorema Si X e Y son variables aleatorias independientes, entonces

$$Cov(X, Y) = 0$$

Teorema Si X e Y son variables aleatorias con función de probabilidad conjunta y varianzas finitas, entonces

$$Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y) \pm 2abCov(X, Y)$$

Definición Sean X e Y variables aleatorias con desviación estándar σ_X y σ_Y respectivamente. El *coeficiente de correlación* de X e Y, denotado $\rho(X, Y)$, se define por

$$\rho(X, Y) \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Teorema Propiedades del coeficiente de correlación ρ

1. $-1 \leq \rho \leq 1$
2. $\rho(X, Y) = \rho(Y, X)$
3. Si X e Y son independientes, entonces $\rho = 0$
4. Si $\rho = \pm 1$, entonces entre X e Y existe una dependencia funcional lineal.
5. Suponga que $Y = aX + b$, donde a y b son constantes.
   Si $a > 0$, entonces $\rho(X, Y) = 1$; si $a < 0$, entonces $\rho(X, Y) = -1$
6. Si a, b, c, d son constantes, $a > 0$ y $b > 0$, entonces $\rho(aX + c, bY + d) = \rho(X, Y)$

1.3. Distribuciones bidimensionales continuas

Definición Si (X, Y) es una variable bidimensional contínua. *La función de densidad de probabilidad conjunta*, es una función integrable $f(x, y)$ que cumple las siguientes propiedades:

1. $f(x, y) > 0$, $\quad \forall (x, y)$
2. $\displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$

La probabilidad de un evento A definido en un plano euclideano, está dado por

$$P[(X, Y) \in A] = \iint_A f(x, y) dx dy$$

Y por tanto $P[(X, Y) \in A]$ es el volumen del sólido sobre la región A es el plano XY y acotado por la superficie $f(x, y)$.
Las *funciones de Densidad Marginal son*:

1. $f_X(x) = \displaystyle\int_{-\infty}^{+\infty} f(x, y) dx$, función densidad marginal para X.
2. $f_Y(y) = \displaystyle\int_{-\infty}^{+\infty} f(x, y) dy$, función densidad marginal para Y.

El valor esperado de X (media de X), el valor esperado de Y (o media de Y) están dados por

$$\begin{aligned} E(X) = \mu_X &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dy dx \\ E(Y) = \mu_Y &= \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dy dx \end{aligned}$$

En general, el valor esperado de una función $H(X, Y)$ se define por

$$E[H(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(x, y) f(x, y) dx dy$$

Las propiedades son las mismas que en el caso discreto, entonces

$$E[H(X, Y) \pm G(X, Y)] = E[H(X, Y)] \pm E[G(X, Y)]$$

La varianza de X e Y están definidos como sigue

$$\begin{aligned} Var(X) = \sigma_X^2 &= \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)^2 f(x, y) dy dx \\ Var(Y) = \sigma_Y^2 &= \int_{-\infty}^{+\infty} (y - \mu_Y)^2 f_Y(y) dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (y - \mu_Y)^2 f(x, y) dy dx \end{aligned}$$

También

$$\sigma_X^2 = E(X^2) - (\mu_X)^2$$

2. Función de distribución binomial y poissoniana

2.1. Ensayo de Bernoulli

Hay muchos experimentos que tienen olo dos resultados posibles, llamados *éxito* E y *fracaso* F. Luego, el espacio muestral para este tipo de ensayo es $\Omega = \{E, F\}$.
Un experimento con esta característica se llama un *ensayo de Bernoulli.* Definimos la variable aleatoria X de tal manera que $X(w) =$ número de exitos en un ensayo de Bernoulli. $R_X = \{0, 1\}$
Es decir: $X(F) = 0$, "si el resultado del ensayo es una F" y $X(E) = 1$, "si el resultado del ensayo es una E".
La variable aleatoria X, así definida se llama, *variable aleatoria de Bernoulli.*
Denotemos $p = P[E]$ y $q = 1 - p = P[F]$.
La distribución de probabilidad de la variable aleatoria X de Bernoulli, es llamada *distribución de Bernoulli*
La distribución de Bernoulli se escribe también como una función así,

$$p(x) = P[X = x] = p^x(1-p)^{1-x},\ x = 0 \text{ ó } x = 1$$

La media y la varianza de la variable aleatoria X, se calcula como sigue:

$$\begin{aligned} \mu &= E(x) = 0.q + 1.p = p \\ \sigma_X^2 &= Var(X) = E(X^2) - (E(X))^2 \\ &= 0^2.q + 1^2.p - p^2 = p - p^2 = p(1-p) \\ &= pq \end{aligned}$$

Hay muchos problemas en la cual el experimento consiste de n ensayos de Bernoulli. se dice que una secuencia de n ensayos iguales de Bernoulli, forman un *proceso de Bernoulli* o *proceso binomial.*

2.2. Distribución binomial

Es el resultado de un experimento aleatorio con las siguientes características:

1. Los resultados posibles son 2: A o $\overline{A}$
2. La probabilidad de A es constante, $P[A] = cte$.
3. Las pruebas son independientes.
4. Debo definir el número total de pruebas.

La *distribución binomial* de la variable aleatoria X es

$$\boxed{P[X = x|B; n, p] = P[X = x] = \binom{n}{x} p^x q^{n-x}\ ,\ x = 0, 1, 2, \cdots, n}$$

Algunas veces, escribiremos simplemente

$$b(x; n.p) = \binom{n}{x} p^x q^{n-x}\ , \qquad x = 0, 1, 2, \cdots, n$$

Observe que:

1. $p(x) = P[X = x] > 0\ , \qquad x = 0, 1, 2, \cdots, n$
2. $\displaystyle\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} P[X = x] = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1^=1$, ya que $p + q = 1$.

   La función de *distribución acumulada* está dada por

$$F(x) = P[X \le x] = P[X \le x|B; n, p] = \begin{cases} 0 & , x < 0 \\ \sum_{k=0}^{[x]} \binom{n}{k} p^k q^{n-k} & , 0 \le x < n \\ 1 & , x \ge n \end{cases}$$

2.2.1. Características de la distribución binomial

La media de la variable aleatoria binomial se puede calcular utilizando directamente la definición.

$$\begin{aligned}
\mu = E(X) &= \sum_{x=0}^{n} x \binom{a}{b} p^x q^{n-x} = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=0}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!} p^x q^{n-x} \\
&= np \sum_{x=0}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}
\end{aligned}$$

Cambiando el índice de la suma, $l = x - 1$, se tiene

$$\begin{aligned}
\mu = E(x) &= np \sum_{x=0}^{n} \frac{(n-1)!}{l!(n-1-l)!} p^{x-1} q^{n-1-l} \\
&= np(p+q)^{n-1}
\end{aligned}$$

Y como $p + q = 1$, se tiene

$$\boxed{\mu = E(X) = np}$$

Una demostración alternativa de la media usando la esperanza matemática es la siguiente:
Puesto que la varible aleatoria X, está definida como el número de éxitos obtenidos en los n ensayos de Bernoulli, entonces

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

donde X_i, $i = 1, 2, ..., n$ es una variable aleatoria con una distribución de probabilidad de Bernoulli. Es decir, X_i está definido por, 1, si el resultado del ensayo es un éxito y 0 si el resultado del experimento es un fracaso. Hemos visto que

$$E(X_i) = p, \qquad i = 1, 2, ..., n$$

Por lo tanto

$$\begin{aligned}
\mu = E(X) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\
&= p + p + \cdots + p = np
\end{aligned}$$

La varianza, se puede determinar aplicando, la propiedad

$$\begin{aligned}
\sigma_X^2 &= E(X^2) - (E(X))^2 \\
E(X^2) &= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} - \sum_{x=0}^{n} n \binom{n}{x} p^x q^{n-x} + \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x q^n - x + \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x q^{n-x} + np
\end{aligned}$$

pero

$$\begin{aligned}
x(x-1)\binom{n}{x} &= x(x-1) \frac{n!}{x!(n-x)!} = \frac{n!}{(x-2)!(n-x)!} \\
&= n(n-1) \frac{(n-2)!}{(x-2)!(n-x)!} = n(n-1) \binom{n-2}{x-2}
\end{aligned}$$

Por lo tanto

$$\begin{aligned}
\sigma_X^2 &= \sum_{x=2}^{n} n(n-1)\binom{n-2}{x-2}p^{x-2}q^{n-x}p^2 + np \\
&= n(n-1)p^2\sum_{n=2}^{n}\binom{n-2}{x-2}p^{x-2}q^{n-x} + np \\
&= n(n-1)p^2\sum_{k=0}^{n}\binom{n-2}{k}p^{k}q^{n-2-k} + np \\
&= n(n-1)p^2(p+q)^{n-2} + np \\
&= n(n-1)p^2 + np
\end{aligned}$$

Luego

$$\begin{aligned}
\sigma_X^2 = Var(X) &= n(n-1)p^2 + np - n^2p^2 \\
&= np^2(n-1-n) + np \\
&= np(1-p) = npq
\end{aligned}$$

$$\boxed{\sigma_X^2 = Var(X) = npq}$$

De donde $\boxed{\sigma = \sqrt{npq}}$ la desviación estándar de X. Para la varianza también una demostración alternativa para el lector qeu conoce el capítulo 4 es la siguiente:
Hemos visto que $\sigma_{X_i}^2 = pq$. Además sabemos que las variables aleatorias X_i son independientes. Entonces

$$\begin{aligned}
\sigma_X^2 = Var(X) &= Var(X_1) + Var(X_2) + \cdots + Var(X_n) \\
&= pq + pq + \cdots + pq = npq
\end{aligned}$$

2.3. Distribución multinomial

Consideremos un experimento ε con espacio muestral Ω con la siguiente característica:

1. Tiene k posibles resultados $E_1, E_2, \cdots, E_k$ mutuamente excluyentes y colectivamente exhaustivos, es decir $E_i \cap E_j = \emptyset,\ i \neq j,\ i,j = 1,2,\cdots,k;\ \bigcup_{i=1}^{k} E_i = \Omega$.
2. $P[E_i] = p_i$, $i = 1,2,\cdots,k$, tal que $\sum_{i=1}^{k} p_i = 1$. En particular, si $k = 2$, el experimento ε es un ensayo de Bernoulli.

Consideremos ahora una secuencia de n ensayos independientes con la caracteristica (1.) y (2.). Es decir cada ensayo tiene k resultado y p_i, $i = 1,2,\cdots,k$ es constante en cada ensayo. Definimos un avariable aleatoria X_i como sigue
$X_i(w) =$ número de veces que el evento E_i ocurre en las n repeticiones de ε. $R_{X_i} = \{0,1,2,\cdots,n\}$
La distribución de probabilidad conjunta de las variables aleatorias $X_1, X_2, \cdots, X_k$ se llama *distribución multinomial*, se define por

$$p(x_1, x_2, \cdots, x_k) = P[X_1 = x_1, \cdots, X_k = x_k] = \frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}$$

donde $\sum_{i=1}^{k} x_i = n$

Observe que $X_1, X_2, \cdots, X_k$ no son variables aleatorias independientes, puesto que $\sum_{i=1}^{k} x_i = n$. Es decir, conocido los valores de $k-1$ variables aleatorias cualesquiera, se conoce la que falta.
La media y la varianza se las variables aleatorias X_i son

$$E(X_i) = np_i\ ; \qquad Var(X_i) = np_i(1-p_i)\ , \qquad i = 1,2,\cdots,k$$

2.4. Distribución de Poisson

2.4.1. Distribución de Poisson como aproximación de la binomial

Mostraremos ahora la distribución de Poisson como un límite de la distribución binomial, con $\lambda = np$. Donde p debe ser suficientemente pequeño y n grande, de tal manera que np permanece casi constante. La distirbución binomial para x éxitos en n ensayos de Bernoulli es

$$P[X = x|B; n, p] = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, \cdots, n$$

Hacemos $\lambda = np$, luego $p = \frac{\lambda}{n}$ y $q = 1 - p = 1 - \frac{\lambda}{n}$
Reemplazando en la fórmula se tiene $P[X = n] = \dfrac{n!}{x!(n-x)!}$