Capítulo 5:

Modelos para variables discretas

Belchin Kostov, Erik Cobo y José A. González

Jordi Cortés, Rosario Peláez, Marta Vilaró, Laura Riba y Nerea Bielsa

Septiembre 2014

Modelos para variables discretas

* Indica tema más avanzado que conviene mirar pero no es crucial para los ejercicios.

Presentación

Ante un caso raro, un clínico veterano, o uno joven que sepa probabilidad condicionada, dirá: "es más probable que sea una presentación curiosa de una patología frecuente, que la típica de una rara". Además, conscientes de la gran variabilidad de los resultados, interpretarán correctamente la broma de que "sano es un paciente no suficientemente visitado". También los buenos gestores de los almacenes de repuestos y de farmacia sabrán que: "si controlas bien el 20% más frecuente de piezas, satisfarán el 80% de clientes". Estas afirmaciones muestran dominio de la probabilidad y de la variabilidad.

En el capítulo anterior aplicamos la probabilidad a eventos, a variables categóricas. Ahora lo haremos a números, a variables de tipo recuento, como el número de crisis previas, en este capítulo; y a variables continuas, como el colesterol o el peso, en el siguiente, Hoy en día, la simulación y la informática permiten estudiar, casi sin restricciones, cualquier situación. Pero conocer unos modelos básicos de probabilidad le ayudará a prever con agilidad muchas situaciones.

La clasificación en variables discretas y continuas será crucial para estudiar su probabilidad. En las primeras, tiene sentido preguntar por probabilidades tanto concretas como acumuladas. Por ejemplo: ¿cuál es la probabilidad de tener 2 hijos? ¿Y la de tener 2 o menos hijos? Pero las variables continuas pueden tomar cualquier valor y la probabilidad de uno concreto se hace insignificante e irrelevante, diremos que 0. No tiene sentido preguntarse por la probabilidad de que alguien pese *exactamente* 70 Kg., pero sí, por ejemplo, que pese entre 69 y 70. O menos de 70.

Pero no se asuste, no deberá aplicar fórmulas ni aprender a usar tablas. R acude en su ayuda y podrá centrarse en su objetivo: aplicar correctamente los resultados y saber cuándo utilizarlos.

Céntrese especialmente en aquellas definiciones que vayan acompañadas de ejercicios. Compruebe que entiende los conceptos de esperanza y varianza. Y que distingue entre valores poblacionales y muestrales.

Contribuciones: BK escribió la versión de septiembre de 2013 a partir de los apuntes de EC y JAG de la asignatura de Probabilidad y Estadística de la Facultad de Informática de la UPC, que fue revisada por JC y RP. MV y LR revisaron la versión de enero de 2104 y NB y EC la de septiembre de 2014.

1. Variable aleatoria. Funciones de probabilidad.

1.1. Variable aleatoria

Una variable aleatoria informa sobre la expectativa de los valores numéricos de un fenómeno incierto. Al ser incierto: (1) debe tener dispersión, es decir, ser variable; y (2) la probabilidad permite expresar la información disponible sobre sus posibles valores.

Nota: Es un razonamiento deductivo, "antes de observar": no hay muestra. Razonamos como si conociéramos TODA la verdad, como si supiéramos qué pasa en la población.

> **Recuerde**
>
> En este tema estudiamos cómo representar poblaciones.

Lectura: Una variable aleatoria se define como el proceso de convertir un individuo o un objeto en un número.

> **Recuerde**
>
> Una **variable aleatoria** es numérica, puede tomar más de un valor y lleva probabilidades asociadas.

Ejemplo 1.1: La duración de una intervención quirúrgica es un proceso incierto, pero no todos los posibles valores de tiempo son igualmente esperables. Conviene disponer de un método para transmitir cuáles son más esperables.

Lectura: La definición de variable se puede generalizar a vectores de números, variables categóricas o los llamados procesos estocásticos o aleatorios.

1.2. Variables discretas frente a continuas

Una variable discreta puede tomar sólo determinados valores; pero una continua, entre dos posibles valores, puede tomar cualquier otro entre ellos.

Nota: En la era digital, el final del proceso de medida siempre es discontinuo: la balanza del cuarto de baño ha pasado de analógica, donde cualquier valor dentro del rango era posible, a digital, con cierto redondeo. Pero la definición de continua hace referencia al concepto, no al resultado de medida. Así, la supervivencia es continua y la trataremos como tal —aunque su valor final se redondee, quizás a meses o días.

En las variables continuas no tiene sentido preguntarse por la probabilidad de un valor exacto, ya que es tan pequeña que es irrelevante. Formalmente sería siempre cero, por lo que requiere un tratamiento matemático especial. Empezaremos por estudiar las discretas.

1.3. Probabilidades en variables discretas

En este caso tiene sentido preguntarse por la probabilidad concreta de un valor y también por su probabilidad acumulada. Ambas se especifican, para todos los valores de la variable, con la ayuda de las llamadas "Función de Probabilidad" y "Función de Distribución".

Definición

La **Función de probabilidad f_X** de una variable discreta X proporciona la probabilidad de cada valor.

Ejemplo 1.2: Cierto equipo, para conseguir la recanalización de un vaso, emplea 1 dispositivo en el 72% de los casos; 2 en un 26%; y 3 en el 2% restante. De forma más abreviada, la Función de probabilidad f_X de la variable número de dispositivos necesarios vale:

$f_X(1) = P(X=1) = 0.72$

$f_X(2) = P(X=2) = 0.26$

$f_X(3) = P(X=3) = 0.02$

Nota muy técnica: Puede 'cerrarse' esta especificación aclarando que los restantes valores tienen probabilidad 0: "Cualquier x diferente de 1, 2 o 3, tiene Función de probabilidad nula", que simbólicamente sería: $\forall x \neq 1, 2, 3;\ f_X(x) = 0$.

La Figura 1.1 muestra la forma de esta Función de probabilidad y que sólo los valores 1, 2 y 3 son posibles en este ejemplo.

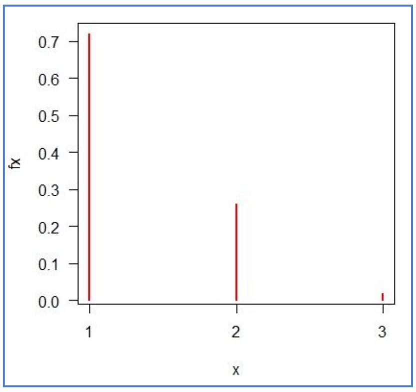


Figura 1.1. Ejemplo de Función de probabilidad

Ejemplo de R

```
# Gráfico de la Figura 1.1
>x <- 1:3
>fx <- c(0.72,0.26,0.02)
>par(las=1)
>plot(x, fx, type = "h", col = 2, lwd = 2,xaxp = c(1,3,2))
```

Ejercicio 1.1

Supongamos que emplear un número bajo de dispositivos sea un objetivo sanitario deseable y se desea definir un indicador que permita al paciente valorar la seguridad que le ofrece una institución. ¿Qué seguridad o confianza tiene el paciente de que el servicio del ejemplo anterior empleará, como mucho 2 dispositivos?

Definición

Llamamos **probabilidad acumulada** de un valor x_i de una variable discreta a la suma de su probabilidad con la de todos los valores inferiores y lo representamos por $P(X \leq x_i)$.

Definición

La **Función de Distribución** F_X de una variable discreta proporciona la probabilidad acumulada para cada valor.

Ejemplo 1.2 (cont.): La Función de Distribución F_X del número X de dispositivos sería:

$$F_X(1) = P(X \leq 1) = 0.72$$

$$F_X(2) = P(X \leq 2) = 0.98$$

$$F_X(3) = P(X \leq 3) = 1$$

Nótese, p.e., que $P(X \leq 1) = P(X \leq 1.5)$.

Nota: También puede 'cerrarse' aclarando los restantes valores: "para x menores de 1, $F_X(x)=0$"; "para x mayores de 3, $F_X(x)=1$"; "si $1 \leq x < 2$, $F_X(x)=0.72$; si $2 \leq x < 3$, $F_X(x)=0.98$".

La Figura 1.2 muestra la forma de su Función de Distribución.

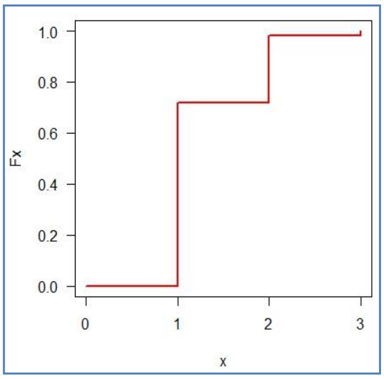


Figura 1.2. Ejemplo de Función de Distribución

Ejemplo de R

```
# Gráfico de la Figura 1.2
> x <- 0:3
> Fx <- c(0,0.72,0.98,1)
> par(las=1)
> plot(x,Fx,type="s",col=2,lwd=2,xaxp=c(0,3,3))
```

Recuerde

La Función de Probabilidad aplica a un valor concreto: $f_X(2) = P(X=2)$

La Función de Distribución acumula probabilidades: $F_X(2) = P(X\leq 2)$

Ejercicio 1.2

Supongamos la variable X número de hijos puede tomar los valores 0, 1, 2, 3, … k. ¿Cuál de las siguientes, a o b, es cierta en cada caso?

1.- a) $P(X=3) = f_X(3)$ b) $P(X=3) = f_X(2)$

2.- a) $P(X=3) = F_X(4) - F_X(3)$ b) $P(X=3) = F_X(3) - F_X(2)$

3.- a) $P(X>3) = 1 - F_X(2)$ b) $P(X>3) = 1 - F_X(3)$

4.- a) $P(X\geq 3) = 1 - F_X(2)$ b) $P(X\geq 3) = 1 - F_X(3)$

Y de forma simbólica, más difícil, pero con la pista de las anteriores:

5.- a) $P(X=k) = f_X(k)$ b) $P(X=k) = f_{k-1}$

6.- a) $P(X=k) = F_X(k+1) - F_X(k)$ b) $P(X=k) = F_X(k) - F_X(k-1)$

7.- a) $P(X>k) = 1 - F_X(k-1)$ b) $P(X>k) = 1 - F_X(k)$

8.- a) $P(X\geq k) = 1 - F_X(k-1)$ b) $P(X\geq k) = 1 - F_X(k)$

2. Esperanza y varianza. Aplicaciones

2.1. Esperanza de una variable, E(X), μ

Calculamos el centro poblacional igual que el muestral, promediando entre todos los valores.

Recuerde

La esperanza E(X)=μ es un parámetro poblacional; el promedio $\bar{x}$, un resultado muestral.

Nota: El promedio se realiza mediante la suma de cada valor ponderado por su probabilidad: $\mu=E(X)=\Sigma_i x_i P(X=x_i)$.

Definición

Llamamos **esperanza** de una variable X a su centro poblacional y lo representamos por E(X) o por μ.

Nota: La fórmula cambia para discretas y continuas, pero en ambos casos pondera por probabilidad.

Interpretamos μ=E(X) como el valor central en el sentido de "centro de gravedad": punto que mantiene "en equilibrio" la distribución de probabilidad.

Ejemplo 1.2 (cont.): La Figura 2.1 vuelve a mostrar la distribución del número X de dispositivos y marca su esperanza: μ=E(X) = 1.3. Puede imaginar que el punto 1.3 coincide con el fiel de una balanza que aguantara, en equilibrio, ambos brazos.

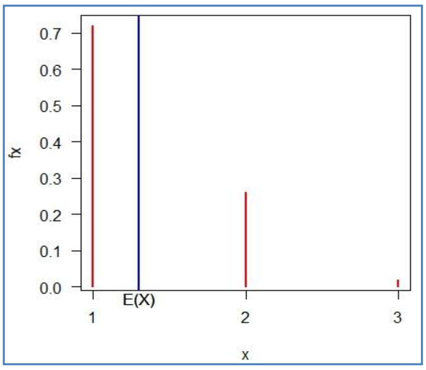


Figura 2.1. Representación del valor esperado, E(X) dentro de f_x

Nota: Aunque no lo parezca, la posición de equilibrio de un cuerpo no deja la misma masa a cada lado: depende de cómo se distribuye esa masa ("ley de la palanca"). Tampoco la esperanza divide la distribución en 50% y 50% (ese punto es la "mediana").

Ejemplo de R

```
# Gráfico de la Figura 2.1. Se añade a la Figura 1.1
> abline(v=1.3,col=4,lwd=2)
> mtext("E(X)",1,at=1.3)
```

Nota: Observe cierto abuso del lenguaje al decir que μ=E(X) es el valor *esperado* de X, ya que no esperamos observar un uso de 1.3 dispositivos en la próxima intervención: veremos 1 o 2 o 3, pero no 1.3.

Recuerde

En una discreta, la esperanza podría NO ser uno de los valores posibles.

E(X)=μ tiene 2 propiedades muy importantes, según la apliquemos a la población o a las unidades.

1) E(X)=μ caracteriza a la población y la diferencia del resto de poblaciones.

Ejemplo 1.2 (cont.): μ = 1.3 es propia de ese entorno: podríamos negar que 1.3 represente a un paciente de esa población. Pero SÍ que representa a esa población: si desde un punto de vista clínico y asistencial es relevante, μ = 1.3 podría ser una medida del rendimiento de esa población.

Recuerde

E(X)=μ es un indicador de la situación de la población.

2) Aplicada a las unidades, E(X)=μ minimiza el error: si "esperamos" E(X)=μ, cometemos cierto error; pero, a lo largo de todos los valores posibles y teniendo en cuenta su frecuencia, será el menor error posible –tal como veremos en predicción en el curso de observacionales.

Ejemplo 1.2 (cont.): Si afirmamos que la intervención futura precisará exactamente 1.3 dispositivos, seguro que NO acertamos; en cambio, hacemos mínimo el error de predicción: la suma de todos los posibles errores es mínima en el sentido de que predecir cualquier otro valor conduce a una suma mayor de errores.

Recuerde

"Esperar" ver E(X)=μ en la próxima observación minimiza el error posible.

Nota técnica: La definición concreta de este error de predicción será 'cuadrática'.

Ejercicio 2.1

Diga para cada una si es cierta. Si fuera falsa, escríbala bien:

a. Tanto la esperanza [E(X) = μ] como el promedio [= media = $\bar{x}$] visto en descriptiva indican cierto centro.

b. La esperanza, E(X) o μ, aplica a los resultados de una muestra; pero el promedio, media o $\bar{x}$, a la distribución poblacional.

2.2. Varianza de una variable, V(X), σ^2

También calcularemos la dispersión poblacional como la muestral, promediando las distancias al cuadrado de cada valor con su esperanza.

Recuerde

La varianza $V(X)=\sigma^2$ es un parámetro poblacional; la varianza S^2, un resultado muestral.

Nota: La fórmula también pondera por la probabilidad de cada valor.

Es interesante expresar la varianza en términos de esperanza:

Definición

Llamamos **varianza** al valor esperado de la distancia cuadrada con la media: $\sigma^2 = V(X) = E\,(X-\mu)^2$.

Como "(X-μ)" es la distancia entre la variable X y su centro, $\sigma^2=V(X)$ es precisamente el valor esperado del error al cuadrado que cometemos al esperar μ cuando observamos X.

Nota: μ era el centro de gravedad de X. Por tanto, si no se elevara al cuadrado, los errores positivos y negativos se compensarían y su suma daría 0.

La varianza está expresada en unidades de X, pero elevadas al cuadrado. Por ello, como en descriptiva, definimos la desviación típica σ como su raíz cuadrada.

Definición

La raíz cuadrada de la varianza se denomina **desviación típica** o estándar y se representa por $\sigma = \sqrt{V(X)}$.

Interpretamos la desviación típica σ como la distancia "esperada" entre las observaciones y su media μ.

Ejercicio 2.2

Behar, Grima y Marco proponen un partido de basket entre marcianos y terrícolas. Suponiendo que sus alturas tengan la misma μ, pero la de los marcianos mayor σ, ¿quién espera que gane si ambos seleccionan a sus jugadores más altos?

Ejemplo 1.2 (cont.): La Figura 2.2 vuelve a mostrar la distribución del número X de dispositivos. Ahora, además de indicar su esperanza: E(X) = 1.3, también marca su desviación típica σ=0.5. Puede imaginar que la distancia promedio de todos los valores de X al punto 1.3 vale 0,5. Su varianza es σ^2=0.25.

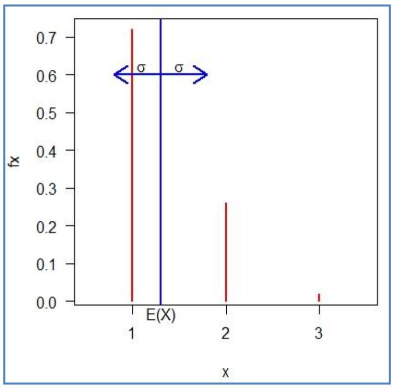


Figura 2.2. Representación gráfica de la desviación típica

Nota: Observe que esta situación es asimétrica y, a la derecha de μ, σ parece quedarse corta, pero a su izquierda, pasarse de largo.

Ejemplo de R

```
# Gráfico de la Figura 2.2
> par(las=1)
> plot(x,fx,type="h",col=2,lwd=2,xaxp=c(1,3,2),
          xlim=c(0.5,3.5))
> abline(v=1.3,col=4,lwd=2)
> mtext("E(X)",1,at=1.3)
> arrows(1.3-0.5, 0.6, 1.3+0.5, 0.6, col=4, lwd=2, code=3)
> text(c(1.1,1.5),0.62,expression(sigma))
```

Sólo en distribuciones simétricas podemos aplicar la desviación típica σ a ambos lados.

Nota: Observe que NO podríamos representar la varianza en ese gráfico, ya que tiene unidades diferentes (son ‘cuadradas’).

Igual que E(X), V(X) o bien su raíz σ también resumen el comportamiento de una población.

Ejemplo 1.2 (cont.): La Figura 2.3 muestra la distribución del número X’ de dispositivos en otra población: ahora han aumentado tanto los aciertos a la primera ($f_{X'}(1)=0.84$) como la necesidad de recurrir a una tercera ($f_{X'}(3)=0.14$). Como resultado, su esperanza, μ=1.3, sigue igual, pero ahora su desviación típica es mayor, σ=0.7. Aunque su rendimiento promedio es el mismo, en global es menos similar: quizás podríamos decir menos igualitaria.

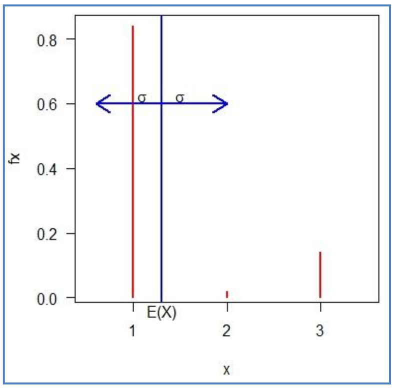


Figura 2.3. Representación gráfica de la desviación típica

Recuerde

V(X) y σ indican la dispersión de la población.

Aplicada a las unidades, σ complementa a μ. E(X)=μ indica el valor central, esperado. Y σ, la distancia esperada a μ.

Ejemplo 1.2 (cont.): X y X’ tienen la misma esperanza, en ambos caos tienen el mismo centro, 1.3 dispositivos. Pero como X’ tiene dispersión mayor, debemos prepararnos para ella.

Ejercicio 2.3

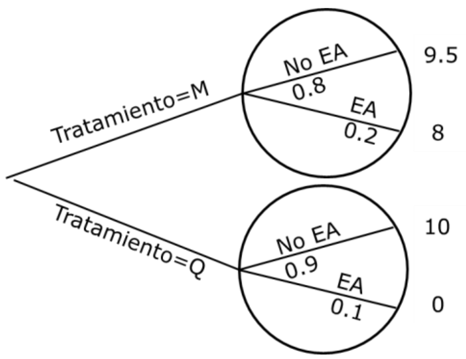
Compare la varianza ($V(X) = \sigma^2$) de este tema con la vista en descriptiva (S^2) y explique sus similutudes y diferencias.

2.3. Aplicaciones. Decisión *

Esperanza y varianza resumen el comportamiento de una población y, por tanto, podemos usarlas para comparar poblaciones. Y para escoger entre ellas. Si cambiamos una intervención A por un B en una población, tendremos 2 posibles poblaciones: aquella que observaríamos cuando aplicamos A y la que observamos al aplicar B.

Ejemplo 2.1: pongamos que una compresión benigna de próstata pueda ser tratada médica (M) o quirúrgicamente (Q). Y que ambas tienen un éxito del 100%, pero difieren en su seguridad: M tiene un 20% de eventos adversos (EA), por un 10% de Q. Sin más información, preferiremos Q. Pero quizás necesitemos profundizar más: ¿Qué pasa en ambos casos si se presenta el EA? Supongamos que la variable de interés es la cantidad de vida en años.

El árbol presenta primero un **nudo** con 2 **opciones** o ramas entre las que elegir: el círculo superior representa la población si optamos por M: la variable tiempo de vida toma el valor 9.5 años en el 80% de las observaciones y 8 años en el 20% restante. Si optamos por Q, habrá otra población en la que el valor 10 años tendrá probabilidad 0.9 y el valor 0 años, 0.1.



Recuerde

Un árbol de decisión contiene **nudos** para escoger entre opciones alternativas y luego variables con sus **probabilidades** para cada **valor**.

Nota: Este simplista ejemplo anula la variabilidad dentro de cada resultado (p.e.: todos los Q sin complicaciones viven 10 años), pero la discusión sería muy parecida si fuera más realista.

Ejemplo 2.1 (cont.): En la población Q, E(X) = 10*0.9+0*0.1 = 9 años. Y en M, E(X)= 9.2. Así, por cada caso en la población, optar por M implica, 0.2 años más de vida por habitante. Globalmente, el conjunto de los 100 habitantes de una pequeña población dispondría de 20 años adicionales.

Recuerde

E(X) valora, como criterio de decisión, el beneficio poblacioal.

Nota: La incertidumbre en Q es mayor que en M. De hecho, sus varianzas son 9 y 0.36 (σ = 3 y 0.6). Así, si quisiéramos disminuir al máximo el grado de "sorpresa" global, escogeríamos el valor mínimo de V(X), que también conduce a M

Nota: ¿Y a nivel individual? Alguien podría decir que quiere evitar, como sea, la peor opción posible. O, según sea su grado de aversión o amor al riesgo, todo lo contrario.

Ejemplo 2.1 (cont.): Cierto paciente recibirá el premio Nobel en unos meses. Sabe que tiene que solucionar el tema o no aguantará toda la ceremonia. Sin duda, querrá evitar Q y la probabilidad de quedarse sin recibir el premio.

Nota: ¿Y la calidad de vida? Por supuesto, elegir otro objetivo requiere recoger otra información, lo que podría llevar a otra decisión. Por eso, lo más importante es "saber qué es lo que uno quiere". Pero marcar cuál es el propio objetivo, conocerse a uno mismo, es quizás lo más difícil de todo. Vea en Wikipedia la pirámide de Maslow.

3. Modelos de probabilidad discretos

Los modelos de probabilidad son situaciones simplificadas que pueden ser útiles.

3.1. Indicador o bernouilli

El más sencillo de todos es el de Bernouilli (B_1) que hace referencia a 1 observación de la variable binaria "indicador", que puede tomar los valores '1', pongamos éxito; y '0', fracaso.

Ejemplo 3.1: Si definimos 'éxito' al evento "eliminar el trombo con el uso de un solo dispositivo", disponemos de una variable B_1 que puede tomar el valor 1 con probabilidad:

$P(X=1) = 0.72 = \pi$

y el valor 0, con probabilidad

$P(X=0) = 0.28 = 1-\pi$

Su esperanza es igual a la probabilidad π de éxito: $E(B_1) = \pi$. Y su varianza es igual al producto de dicha probabilidad por su complementario: $V(B_1) = \pi\cdot(1-\pi)$.

Ejercicio 3.1

Según la variabilidad de la Bernouilli, ¿qué es más incierto, un indicador con π=0.5 o con π=0.1? ¿Cuadra este resultado teórico con su intuición previa de incierto para valores de π=0.5 o de π=0.1? ¿Se atreve a decir qué valor de π va acompañado de mayor incertidumbre?

3.2. Binomial

Si repetimos 'n' veces el indicador de Bernouilli anterior y contamos el número de éxitos, aparece el modelo Binomial (B_n) siempre que se cumplan las 2 condiciones siguientes: (1) las 'n' repeticiones tienen todas la misma probabilidad π de éxito; y (2) todas ellas son mutuamente independientes.

Definición

Representamos por $B(n,\pi)$ a la variable discreta recuento de éxitos tras 'n' repeticiones independientes de indicadores B_1 con probabilidad de éxito π.

Ejemplo 3.2: Siguiendo con el ejemplo anterior, si un servicio hace 10 intervenciones a la semana (n=10), el número de éxitos se modela con una distribución B(10, 0.72).

3.2.1. Función de Probabilidad f_X y función de Distribución F_X

Recordemos que hay que distinguir entre la Función de Probabilidad, f_X, que corresponde a un valor exacto, y la Función de Distribución, F_X, que acumula la probabilidad de los valores previos. La siguiente figura muestra las funciones de probabilidad f_X y de distribución F_X del ejemplo 3.2. Veamos cómo obtenerlas con R.

Instrucciones en R para fx y Fx

```
#  fx: Probabilidad de obtener un 2 en una B(n=6,Pi=2/3)
> dbinom(2, 6, 2/3)
[1] 0.08230453
#  Fx: Probabilidad de obtener un 2 o menos en una B(n=6,Pi=2/3)
> pbinom(2, 6, 2/3)
[1] 0.1001372
#  1-Fx: Probabilidad de obtener un 3 o más en una B(n=6,Pi=2/3)
> 1-pbinom(2, 6, 2/3)
[1] 0.8998628
```

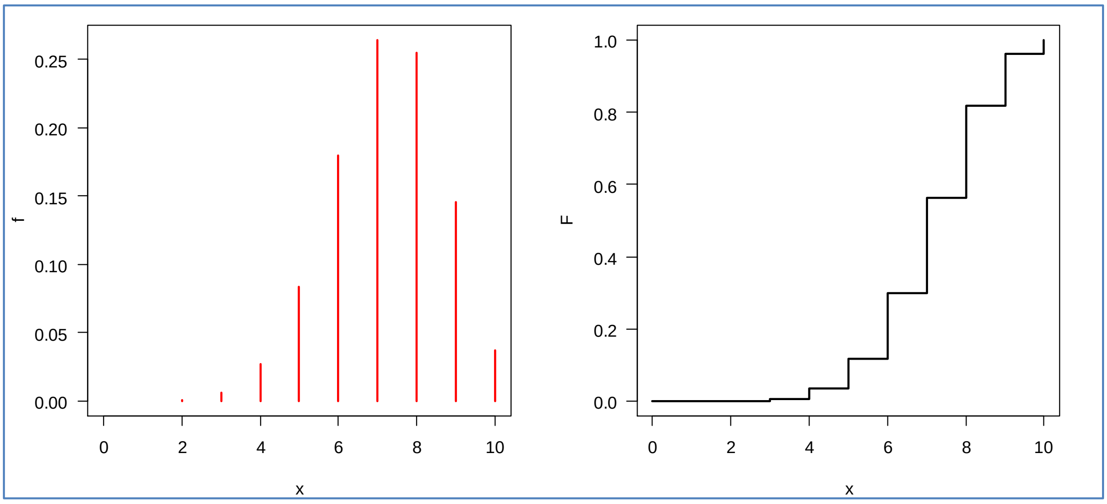


Figura 3.1. Funciones de probabilidad f_X y de distribución F_X de una B(10, 0.72)

```
#Instrucciones de la Figura
par(mfrow=c(1,2), las=1)
x=0:10
f = dbinom(x, 10, 0.72)
F = pbinom(x, 10, 0.72)
plot(x, f, t='h', lwd=2, col='red')
plot(x, F, t='s', lwd=2)
```

Ejercicio 3.2

a) ¿Bajo qué premisas serían ciertos los cálculos anteriores?

b) ¿Le parecen razonables?

Notación:

Se indica que X se modela con una B(n,π) mediante: X~B(n,π)

Ejemplo de R

```
# Cálculo de fx: P(X=7) si X~B(10,0.72)

>dbinom(7,10,0.72)
[1] 0.2642304
```

Ejercicio 3.3

Interprete el resultado anterior: ¿le parece bajo o alto que en 10 observaciones, la probabilidad de observar 7 éxitos sea aproximadamente de ¼, siendo π=0.7?

Ejercicio 3.4

Calcule con R la probabilidad de observar 8 y de observar 6. ¿Son mayores o menores que las de 7? Interprete.

Ejemplo de R

```
# Cálculo de Fx: P(X≤7) si X~B(10,0.72)
> pbinom(7,10,0.72)
[1] 0.562171
```

Ejercicio 3.5

Calcule con R la probabilidad de observar 8 o menos. Y la de observar 5 o menos. Deduzca la probabilidad de observar 6, 7 u 8: $P(6 \leq X \leq 8)$. Interprete.

Ejercicio 3.6 *

Imagine ahora que los resultados se presentan por trimestres, donde se realizan 100 en lugar de 10 intervenciones, con la misma probabilidad π=0.72. Calcule con R $P(X=70)$, $P(X \leq 70)$ y $P(60 \leq X \leq 80)$. Compare con los resultados anteriores teniendo en cuenta que la proporción observada es la misma en 7 de 10 y en 70 de 100. Interprete.

Recuerde

El modelo Binomial estudia la probabilidad de observar X eventos en n repeticiones de un indicador con probabilidad π.

3.2.2. Esperanza y varianza

Las expresiones de la esperanza y la varianza de una variable Binomial (B_n) son:

Fórmulas

Si $X \sim B(n,\pi)$, $E(X) = n \cdot \pi$ y $V(X) = n \cdot \pi \cdot (1-\pi)$

Ejercicio 3.7

Calcule E(X) y V(X) para $X_{10} \sim B(10,0.72)$ y para $X_{100} \sim B(100,0.72)$. Interprete ambas. [Si hizo el ejercicio reto anterior, compare resultados].

En una Binomial, la definición de éxito y fracaso es pura convención. Imaginemos que lo que preocupa es justamente lo contrario: que no se consiga la revascularización con el primer dispositivo y queremos contar el número de veces que no se consigue.

Ejercicio 3.8

Especifique la distribución de la variable Y_{10}: número de fracasos semanales (10 intervenciones). Ídem para Y_{100} (trimestrales). Calcule $P(Y_{10} \leq 1)$ y la de $P(Y_{100} \leq 10)$.

Ejemplo 3.3: Suponga que una cierta analítica consta de 20 pruebas, cada una de ellas con una especificidad del 95%, es decir, el 95% de los sanos da negativo en cada una de las pruebas. Suponga también que son independientes entre sí, es decir que valoran entidades diferentes y que el hecho de dar positivo una de ellas no aumenta la probabilidad de que otra sea positiva. Es decir, que podemos modelar el número de resultados positivos (X) por una $B(n=20,\pi=0.05)$. El número de resultados que cabe resultar que sean positivos es algo preocupante: $E(X)=n \cdot \pi=20 \cdot 0.05=1$. Pero mucho más preocupante es la probabilidad de que un sano dé negativo en todos ellos: $P(X=0)=$ dnorm(0,20,0.05)= 0.3584859. Es decir, sólo un 36% de los sanos darán negativo en todas las pruebas. O peor aún ¡es más probable que un sano tenga algún resultado positivo que todos negativos.

Historieta: ”Sano es alguien no suficientemente visitado por un médico”.

Historieta: ”Tanto va el cántaro a la fuente que al final se rompe”.

Recuerde

Valore con prudencia los resultados positivos inesperados tras la repetición de pruebas con especificidad por debajo del 100%.

3.2.3. Cuantiles

Hasta ahora hemos aprendido a calcular probabilidades acumuladas a partir de los valores de X. Pero podríamos tener justo el interés contrario: dada una probabilidad acumulada deseada, conocer cuál es el valor que la proporciona.

Recuerde

Hay 2 tipos de problemas: saber el valor de X y buscar cierta probabilidad; o conocer la probabilidad y buscar el valor de X.

Ejemplo 3.4: ¿Cuántos caben, sin doblar las piernas, en camas de 1.80 metros? Si queremos que quepan un 99%, ¿qué longitud debe tener la cama?

Recuerde

En el tema 2 de descriptiva, al hablar de la mediana y los percentiles, definimos los **cuantiles** como las proporciones muestrales acumuladas. También usaremos **cuantil**, a nivel poblacional, para las probabilidades acumuladas.

Recuerde

Los cuantiles de uso más frecuente son los percentiles, los cuartiles y los quintiles.

Ejemplo 3.5: Siguiendo con el ejemplo de los dispositivos y los resultados semanales con 10 intervenciones, interesa calcular el número máximo de fracasos que podemos garantizar en el 90% de las semanas. Sabiendo que $P(Y_{10} \leq 4) = 0.882$ [pbinom(4,10,0.28)] y que $P(Y_{10} \leq 5) = 0.966$ [pbinom(5,10,0.28)], el valor que garantiza una confianza del 90% es 5. Se trata pues del percentil 0.90. En resumen, podemos garantizar que en el 90% de las semanas como mucho 5 pacientes precisarán más de una intervención. Al ser discretas, "salta" de 0.882 a 0.966, por lo que el percentil 90 coincide, p.e., con el percentil 95.

Recuerde

En las discretas, los cuantiles también dan saltos. Para garantizar una cierta probabilidad se toma el valor superior de la variable.

Ejercicio 3.9

Pasemos a los resultados trimestrales con 100 pacientes. Sabiendo que $P(Y_{100} \leq 33)=0.888$ y $P(Y_{100} \leq 34)=0.924$. Calcule el percentil 90. Interprete. Ofrezca una garantía del 90% sobre el rendimiento del servicio.

Ejemplo de R

```
# Cuantil para Y~B(100,0.72)
# Valor k que cumple 0.9 = Fk = P(Y≤k)
> qbinom(0.90,100,0.72)
[1] 34
```

Ejercicio 3.10

Obtenga los percentiles 0.95 de $Y_{10} \sim B(10,0.28)$ y $Y_{100} \sim B(100,0.28)$. Interprete.

Ejercicio 3.11

Suponga que está investigando una nueva intervención. Vd. recuerda los casos de la talidomida y del TGN1412, dos productos de nueva creación, uno químico y otro biológico, que tuvieron eventos adversos (EA) muy graves. Vd. está contento porque no ha observado ningún EA grave en 10 pacientes. Y Vd. considera que, dados sus grandes efectos positivos, el producto aún será útil incluso si la probabilidad de un EA grave alcanza el 10% ($\pi=0.1$). Su recogida de información le permite descontar la posibilidad de contagios y puede asumir la independencia que requiere la Binomial. Bajo este modelo, si el producto tuviera una $\pi=0.1$ común para todos los pacientes, ¿cuál sería la probabilidad de obtener 0 de 10 casos con un EA grave? Interprete. Repita para un resultado hipotético de 0 sobre 100. Interprete.

3.3. Poisson

El modelo de Poisson es un caso particular del Binomial especialmente útil cuando es más factible obtener el valor del producto $\pi \cdot n$ que los valores exactos de π y de n.

Ejemplo 3.6: El recuento del número diario de accidentados con lesiones craneoencefálicas que requieren un neurocirujano de urgencias es una variable que puede aproximarse por una Binomial. En cierta población es factible conocer la esperanza de esta variable, pero 'n' es tan grande y (afortunadamente) π es tan baja, que resulta complicado obtener n y π con precisión. En cambio, podemos conocer cuál es el promedio en el pasado de este número.

Ejercicio 3.12

Vamos a jugar un poco. Vamos a ver qué pasa con la probabilidad de observar 0 casos si cambiamos π y n de forma que mantengamos constante su producto $\pi \cdot n$. Recupere del ejemplo anterior P(X=0) para B_{10} (10, 0.1) y obténga también P(X=0) para B_{100}(100, 0.01), B_{1000}(1000, 0.001) y $B_{1000000}$ (1000000, 0.000001). Interprete. Calcule la esperanza de estas 4 variables. Interprete.

Notación

El producto $\pi \cdot n$ se llama *tasa* y se representa con la letra λ.

Expresamos el modelo de Poisson de tasa (o parámetro) λ mediante P(λ).

Ejemplo 3.7: En Barcelona quizás podríamos tener una esperanza de 1 caso diario. Es decir, de 1 caso por día, donde este 'por' indica división: '/'.

Nota: Hemos dicho 'diario': hay que expresar el periodo.

Notación

La tasa λ *suele* indicar casos/tiempo (vea el tema 4).

Debe, por tanto, especificar el periodo de observación o seguimiento.

Nota: λ es un número (real) positivo que representa la tasa media de casos por lapso de seguimiento considerado. En general, se mide en tiempo (10 casos/semana, por ejemplo).

Ejemplo 3.8: Barcelona tiene en promedio 1 traumatismo craneoencefálico diario (o por día, o "1 evento/día").

Nota: Otros indicadores del denominador o nivel de exposición pueden ser el número de tomas (en el caso del riesgo de un fármaco); o el número de pernoctaciones en un centro sanitario (en el caso de un riesgo de infección). [Si permite un ejemplo cotidiano en el límite de lo correcto, ciertos pseudo-hoteles, podrían tener un promedio de pernoctaciones (o alquileres por día) de λ=5.3]

Ejemplo 3.9: El fármaco tal tiene 1 evento adverso grave cada 1000 tomas.

Ejemplo 3.10: El centro tal tiene 1 infección nosocomial por 1000 estancias.

3.3.1. Funciones de Probabilidad, f_X, y de Distribución, F_X

Ejemplo de R

```
# Cálculo de fx: P(X=0) si X~P(1)
> dpois(0,1)
[1] 0.3678794
```

Ejercicio 3.13

¿Se parece este resultado a los del ejercicio anterior? ¿A cuáles se parece más?

Ejercicio 3.14

Si la tasa diaria de traumatismos craneoencefálicos vale 1, ¿ qué probabilidadtienen 0, 1, 2, 3 y 4 traumatismos? ¿En qué proporción de días se observarán 0, 1, 2, 3 y 4 traumatismos?

Ejemplo de R

```
# Cálculo de Fx: P(X≤2) si X~P(1)
> ppois(2,1)
[1] 0.9196986
```

Ejercicio 3.15

Calcule con R las probabilidades de observar 3 o menos traumatismos. Si Vd. dimensiona sus servicios para atender hasta 4, ¿qué garantías tiene de que un día concreto cubra todas las necesidades?

Recuerde

El modelo de Poisson estudia la probabilidad de observar X eventos por unidad de tiempo cuando su frecuencia de aparición es λ.

3.3.2. Esperanza y varianza

Las expresiones de la esperanza y la varianza de una variable Poisson (P_λ) son:

Fórmulas

Si $X \sim P(\lambda)$, $E(X) = \lambda$ y $V(X) = \lambda$

Ejercicio 3.16

En el modelo de Poisson, $E(X) = V(X) = \lambda$. ¿Cuánto vale V(X) en el caso de los traumatistos? ¿Y σ? Repita para $\lambda=4$. ¿Tiene sentido que la dispersión sea mayor cuanto mayor sea el centro?

Recuerde

En Poisson, parámetro tasa λ, esperanza $E(X)=\mu$ y varianza $V(X)=\sigma^2$ son iguales.

3.3.3. Cuantiles

Como en la Binomial, también podemos calcular los cuantiles para responder la pregunta inversa.

Ejemplo 3.11: En Barcelona hay un promedio de 4 ictus semanales susceptibles de ser tratados con endoscopia vascular y consumimos 1 dispositivo por paciente. Si deseamos disponer en el almacén con dispositivos suficientes para cubrir el 99.9% de las semanas, ¿cuántos dispositivos necesitamos?

Ejemplo de R

```
> qpois (0.999, 4)
[1] 11

# Es decir, el 99.9% de las semanas hay como mucho 11 casos de este tipo. O también: la probabilidad de que en 1 semana concreta tengamos 11 eventos o menos es 0.999. Con 11 dispositivos cubriremos el 99.9% de las semanas.
```

Ejercicio 3.17

Siguiendo con el caso de los traumatismos craneoencefálicos, si Vd. desea que sus servicios estén preparados para atender todas las urgencias el 99% de los días, ¿para cuántos casos han de estar preparados?

Nota: El modelo Poisson es más exacto si n crece, pero $n\pi$ permanece fijo. De acuerdo con dos reglas de oro, esta aproximación es buena si $n \geq 20$ y $\pi \leq 0.05$, o si $n \geq 100$ y $n \cdot \pi \leq 10$.

3.3.4. Premisas *

La premisa más importante del modelo de Poisson consiste en asumir que λ es constante para las unidades consideradas y que los eventos son independientes entre sí: que observar 1 caso no altera las probabilidades de observar otro. Se dice que el proceso en estudio no tiene memoria.

Ejemplo 3.12: Todos los días tienen la misma frecuencia de traumatismos craneoencefálicos. Una vez observado un caso, no aumenta ni disminuye la probabilidad de observar otro.

Ejemplo 3.13: No haber tenido ningún evento adverso grave hasta la fecha no cambia su expectativa futura.

Ejemplo 3.14: Haber observado 1 infección nosocomial en una estancia no cambia la probabilidad de observarla en otras estancias.

Ejercicio 3.18

A) ¿Le parecen razonables las premisas de la Poisson en estos 3 ejemplos?

B) Si no lo fueran, ¿qué sería incorrecto, el valor observado de la tasa λ o los cálculos que obtendríamos con el modelo de Poisson?

C) ¿Cómo cree que podría comprobarlo?

Es posible comparar las frecuencias predichas por el modelo de Poisson con los resultados observados empíricamente. Cuánto más se parezcan, más creíbles serán las premisas en las que se basa dicho modelo.

Ejemplo 3.15 Aberdein y Spiegelhalter observaron una media de 0.6 ciclistas muertos en Londres cada 2 semanas. Como disponían de datos desde 2005 hasta 2012, pudieron contar cuantos periodos de 2 semanas tuvieron 0 eventos, cuántos 1, etc. Las figuras 3.2 y 3.3 muestran muy buen ajuste entre las probabilidades predichas por el modelo y las observadas.

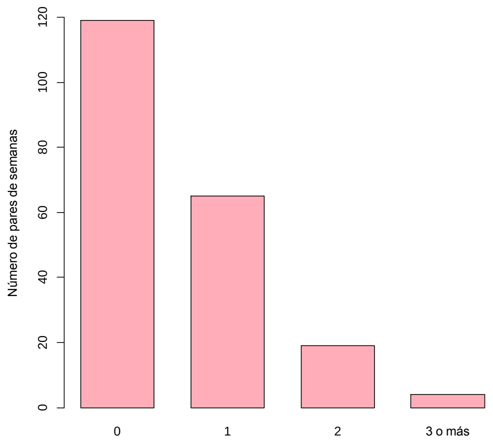


Figura 3.2: muertes esperadas

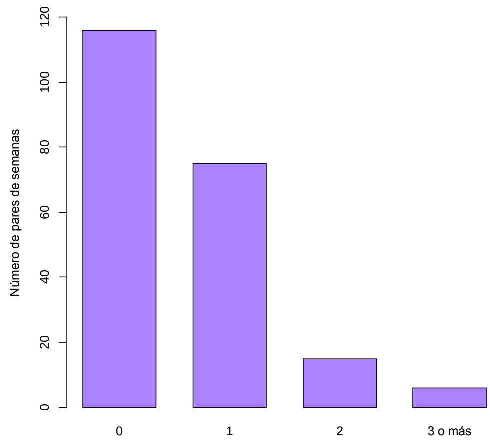


Figura 3.3: muertes reales

En ocasiones, el ajuste entre las predicciones realizadas por el modelo teórico (quizás simple) y los datos observados es muy bueno.

Recuerde

Sea prudente y hable de *modelos* en lugar de *leyes*.

3.3.5. Similitud entre Binomial y Poisson *

Dijimos que el modelo de Poisson es un caso particular del Binomial cuando n crece y π decrece, ambos indefinidamente, pero su producto λ permanece constante.

Nota: La 'n' de Poisson era el número de soldados del ejército de Napoleón y, cómo el matemático Poisson no disponía de R, propuso su modelo para ahorrar tiempo de cálculo.

Los dos siguientes ejemplos muestran que la similitud entre Poisson y Binomial es mayor cuanto mayor es n.

Ejemplo 3.16: La Figura 3.4a muestra la Función de Distribución de dos variables con la misma esperanza: una B(20, 0.5) y una P(10). Puede verse que al inicio crece más rápido P, pero luego B. La discrepancia máxima se observa para x=7, ya que la probabilidad acumulada para P es casi un 9% mayor que para B. En el primer caso, $P[X \leq 7|X \sim P(10)]$ =

0.220 y en el segundo, (P[X≤7|X~B(20, 0.5)] = 0.132. La siguiente mayor discrepancia es para X=13, casi un 8% mayor para B.

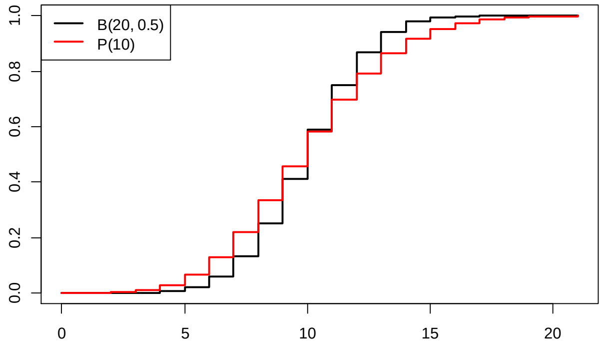


Figura 3.4a. Discrepancia entre Poisson y Binomial de misma esperanza para n pequeña

Ejemplo 3.17: Ahora repetimos el estudio de similitud para la misma P(10), pero con una B(100, 0.1). Se aprecia que el parecido es mucho mejor. Otra vez crece más primero P y luego B. También ahora, el desajuste máximo es para X=7, pero ahora vale 0.014, un 1.4% ya que, para la Binomial (P[X≤7|X~B(100, 0.1)] = 0.206, mucho más cerca de la Poisson (que es la misma). Para X=13, la diferencia es 1.2%. Así, el ajuste es mucho mejor, ya que la mayor discrepancia ha bajado de 8.9 a 1.4.

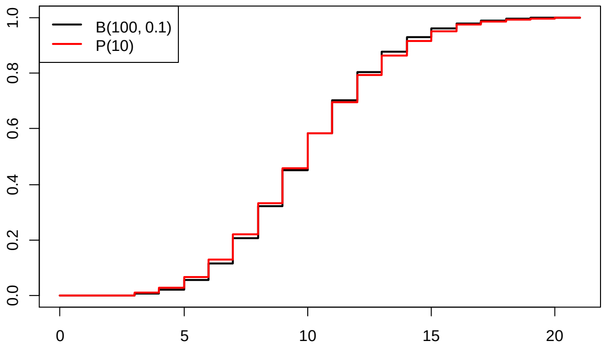


Figura 3.4b. Parecido entre Binomial y Poisson de misma esperanza para n grande.

3.4. Otros modelos *

El modelo de Pascal o Geométrico también es como el binomial, pero en lugar de sumar el número de éxitos, cuenta el número de *fallos previos* al primer éxito.

Nota: Como la Binomial y la Poisson, la geométrica también asume que la probabilidad es siempre la misma e independiente de los resultados previos.

Notación

G(π) representa la variable discreta número de fallos antes del primer éxito.

Ejemplo de R

```
# Recuerde que ??geom le lleva a las instrucciones
# Cálculo de fx: P(X=2) si X~G(0.8)
> dgeom (2,0.8)
[1] 0.032
# Cálculo de Fx: P(X≤7) si X~G(0.2)
> pgeom(7,0.2)
[1] 0.8322278
# Cuantil: k que cumple 0.9 = P(Y≤k) si Y~G(0.3)
> qgeom(0.90,0.3)
[1] 6
```

Ejercicio 3.19

Vamos lanzando una moneda hasta observar una cara. ¿Cuál es la probabilidad de que el número de cruces previas sea 0? ¿Y de que sea 1? ¿Cuál es la probabilidad de observar una cara, como muy tarde, en 5 intentos? [Pista: eso implica 4 fallos previos.]

Ejercicio 3.20 Para obtener cierta acreditación necesitamos 1 artículo en una revista buena (definida como primer cuartil según el factor de impacto). Si un profesional al que le aceptan (de forma independiente) un 15% de artículos, va enviando articulos, ¿qué probabilidad tiene de conseguir la acreditación cuando llegue al décimo (ni antes ni después)? ¿qué probabilidad tiene de conseguir la acreditación al décimo o antes? Interprete. [Pista: si lo debe conseguir al intento número 10, necesita 9 fallos previos.]

Ejercicio 3.21

Suponga que la probabilidad de adquirir una infección nosocomial es la misma en el primer día de estancia que en cualquier día siguiente y que vale un 5%. ¿Cuál es la probabilidad de que un paciente permanenezca ingresado 10 días sin adquirirla? ¿Y uno ingresado 20 días? Suponga ahora que en un paciente más grave esta probabilidad vale un 20%, y repita los cálculos para 10 y 20 días. Interprete: ¿es el coste la única razón para abreviar la estancia hospitalaria?

Nota: Algunos textos definen la geométrica incluyendo también el éxito final.

La Binomial Negativa generaliza la Geométrica al número de fallos antes de alcanzar un número concreto de r éxitos –en lugar de hasta el primer éxito, hasta el r-ésimo éxito.

Ejemplo 3.18: Suponga que al profesional anterior le piden 5 artículos en el primer cuartil en lugar de 1.

Nota: La Poisson tiene la restricción de que la varianza es igual a la esperanza [V(X) = E(X)]. Para liberar esta condición se puede substituir la Poisson por una binomial negativa (BN). La figura 3.5 muestra 3 BN con la misma esperanza que la Poisson pero con una dispersión mayor.

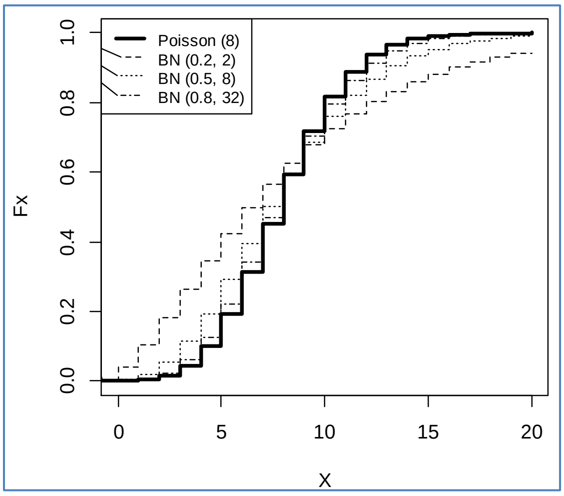


Figura 3.5. 3 BN y 1 Poisson de misma esperanza y distinta varianza

En una población con dos tipos de individuos A y B, la Hipergeométrica cuenta el número de individuos de uno de los tipos al escoger al azar una cantidad determinada de individuos de la población original.

Ejemplo 3.19:Los participantes de cierto estudio clínico son 35 de raza caucásica y 15 de otras razas. Para realizar cierta comprobación se escogen al azar a 10 participantes. ¿Cuál es la probabilidad de que en este pequeño grupo haya como máximo una persona de raza no caucásica? La variable X que cuenta el número de personas no caucásicas escogidas sigue una distribución HGeo(m, n, k) donde m=15, n=35 y k=10, y queremos FX(1): phyper(1,15,35,10) = 0.1209752. La probabilidad de que en este grupo haya como máximo una persona de raza no caucásica es del 12.1%.

Si Vd. dispone de una 'n' muy grande y de las frecuencias observadas para todos los recuentos, puede ahorrarse *imponer* un modelo de probabilidad y trabajar con los resultados observados.

Ejemplo 3.20: En los 47 años que llevamos recogiendo datos, el 45% de los días ha habido 0 intervenciones por traumatismos craneoencefálicos; el 30%, 1; el 16%, 2; el 2%, 3; el 3%, 4 y el 4%, 5.

3.5. Verosimilitud *

Hasta ahora hemos usado estos modelos para, dado un valor del parámetro, calcular las probabilidades de observar ciertos resultados. Pero estos modelos pueden usarse al revés: habiendo observado un cierto resultado, ¿qué valores del parámetro son razonables?

Ejemplo 3.21: Vd. ha observado 6 caras tras lanzar 10 veces la moneda. Si acepta el modelo Binomial, puede calcular la probabilidad de observar esta muestra bajo diferentes valores del parámetro.

Ejercicio 3.22

En el modelo Binomial, ¿cuál es la probabilidad de observar 6 caras de 10 lanzamiento si π=0.6? ¿Y si vale 0.5?

Ejemplo 3.21 (cont.): La Figura 3.6 representa las probabilidades de observar 6 caras de 10 lanzamientos para los valores del parámetro de la Binomial comprendidos entre $0<\pi<1$. Observe que el valor del parámetro para el que la verosimilitud de la muestra es mayor es, precisamente, 0.6. Note también que la probabilidad de esta muestra no es muy grande (0.25), ni cambia demasiado para otros valores muy próximos a 0.6, pero sí al alejarse.

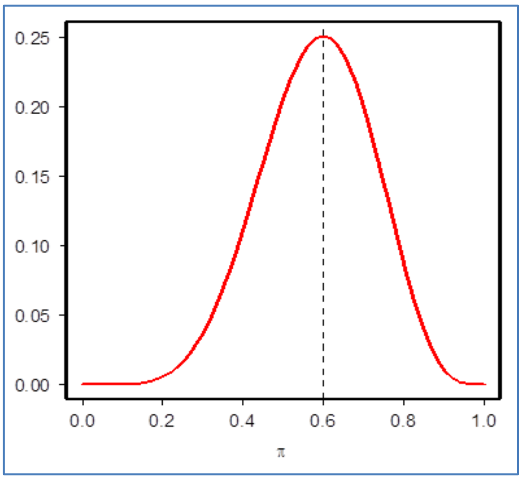


Figura 3.6. Binomial de parámetro $0<\pi<1$

Nota: Si π fuera 1, sólo se podría observar 10 caras en 10 lanzamientos. Si se observan 6 caras de 10, se pueden ya descartar valores de π igual a 0 o 1.

Tiene sentido estimar el parámetro con aquel valor más verosímil.

Recuerde

La verosimilitud se usa para estimar parámetros.

Probabilidad aplica a muestras; verosimilitud, a parámetros.

Ejemplo de R

```
# Función de verosimilitud de π en una binomial si # se observan 60 éxitos en 100 intentos

> x = seq(0,1,len=100)
> curve(dbinom(60,100,x))

# Función de verosimilitud de π en una geométrica  # si se observan 2 fallos antes del primer éxito

> curve(dgeom(2,x))

# Función de verosimilitud de π en una BN si se    # observan 4 fallos antes del tercer éxito

> curve(dnbinom(4,3,x))
```

Soluciones a los ejercicios

1.1 Dado que X es una variable discreta positiva y teniendo en cuenta que la Función de Distribución sólo está definida para los valores X={1,2,3} y que por lo tanto en el resto la probabilidad es 0:

$P(X \leq 2) = P(X=1) + P(X=2) = 0.72 + 0.26 = 0.98$

O bien: $P(X \leq 2) = 1 - P(X>2) = 1 - P(X=3) = 1 - 0.02 = 0.98$

Es decir, el 98% de los casos se atenderán con, como mucho, 2 dispositivos.

1.2 1. a); 2. b); 3. b); 4. a); 5. a); 6. b); 7. b); 8. a)

2.1. a) Cierto.

b) Falso. El promedio, media o $\bar{x}$, aplica a los resultados de una muestra; la esperanza, E(X) o μ, a la distribución poblacional.

2.2. En los marcianos hay más diferencias entre sus alturas al tener una mayor σ, por lo que habrá marcianos con alturas más extremas, tanto más altos como más bajos. Si ambos equipos seleccionan a lo más altos, el equipo de marcianos tendrá jugadores más altos que en el equipo terrícola. Así que el equipo de los marcianos es el favorito —asumiendo igualdad en el resto de factores. [Note que si los jugadores se hubieran seleccionado al azar, esto no pasaría. Eliminar el azar en un proceso de selección suele llevar sorpresas. Veremos más ejemplos de sesgo de selección en el curso de observacionales.]

2.3. La variancia poblacional σ^2 y la muestral S^2 se basan en el mismo principio: medir un promedio de las distancias al cuadrado de los valores al respectivo centro y, por tanto, disponer de un indicador para cuantificar la dispersión de los valores. La diferencia fundamental está en que V(X) lo hace para todos los valores (que potencialmente podrían observarse o no) en base a unas probabilidades conocidas; mientras que S^2 se basa en los valores que se han observado.

3.1. Cuanto mayor es la varianza más incierto es el indicador.

$V(\pi=0.5) = 0.5(1-0.5) = 0.25$

$V(\pi=0.1) = 0.1(1-0.1) = 0.09$

Por lo tanto hay más incertidumbre con $\pi=0.5$.

3.2. a) Las premisas que deberían cumplirse son que las 10 intervenciones tienen la misma probabilidad de éxito (72%), es decir, todos los pacientes provienen de una misma población con los mismos factores de riesgo. Y que las intervenciones son independientes entre sí, es decir que el fracaso o éxito de una no *condiciona* la siguiente.

b) Parece razonable pensar que, si todos los pacientes vienen de la misma población, en cada intervención tenemos la misma probabilidad de éxito. Para poder suponer que el resultado de una intervención no influya en una intervención futura, quizá sea necesario que el equipo vaya cambiando o que sean inmunes al desaliento.

3.3. Como la probabilidad de la binomial es 0.72, que la probabilidad de observar 7 de 10 sea “solo” 0.264 parece un número bajo. Aunque 7 es el valor más probable de observar, observar cualquier otros es más probable que observar un 7: la suma de las probabilidades de obtener un número diferente de 7, 0.736, es mucho mayor.

3.4. Los resultados de R son:

```
>dbinom(8,10,0.72)
[1] 0.2547936
>dbinom(6,10,0.72)
[1] 0.1798235
```

Parece que la probabilidad de obtener un número de éxitos determinado disminuye a medida que este número se aleja del valor esperado ($E=n\cdot\pi=7.2$).

3.5. La probabilidad de observar 8 o menos $P(X\leq 8)$ es:

```
>pbinom(8,10,0.72)
[1] 0.8169646
```

La probabilidad de observar 5 o menos $P(X\leq 5)$es:

```
>pbinom(5,10,0.72)
[1] 0.1181171
```

A partir de las dos probabilidades anteriores podemos calcular la probabilidad de observar 6, 7 u 8:

$P(6\leq X\leq 8) = P(X\leq 8) - P(X\leq 5) = 0.698847$

En 10 intentos con una probabilidad de éxito de 0.72, 7 de cada 10 veces obtendremos un número de éxitos comprendido entre 6 y 8. Tenemos cierta confianza (prob=0.699) de que, al obtener 10 observaciones de una binomial con pi=0.72, el resultado estará cerca de su esperanza, entre 6 y 8).

3.6. En este caso, la n se ha multiplicado por 10, y también los valores de los que queremos hallar las probabilidades (60, 70 y 80, en vez de 6, 7 y 8): aparentemente, las cosas no deberían ser muy distintas. Sin embargo:

```
>dbinom(70,100,0.72)
[1] 0.07869629
```

La probabilidad de observar 70 de 100 (7.9%) es muy inferior a la de observar 7 de 10 (26.4%).

```
>pbinom(70,100,0.72)
[1] 0.3637841
```

La probabilidad de observar menos de 70 sobre 100 también ha bajado: 36% en vez de 56%.

```
>pbinom(80,100,0.72)-pbinom(59,100,0.72)
[1] 0.9706188
```

97% en vez de 70%. Es decir, la probabilidad de valores extremos, más de 80 sobre 100, es del 3%, en lugar del 30% para más de 8 sobre 10. Al aumentar la información, el número de casos disponible, bajan las probabilidades de resultados extremos.

3.7. $E(X_{10})$ = n p = 10 0.72 = 7.2

$V(X_{10})$ = n p (1-p) = 10 0.72 0.28 = 2.016

$E(X_{100})$ = n p = 100 0.72 = 72

$V(X_{100})$ = n p (1-p) = 100 0.72 0.28 = 20.16

Tanto esperanza como variancia se han multiplicado por 10, pero recordemos que el indicador que nos mide la dispersión en unas unidades comparables es la desviación típica, así que mientras el centro se multiplica por 10, la dispersión solo ha aumentado unas 3 veces (3.1623=$\sqrt{10}$).

3.8. $Y_{10} \sim B(10,0.28)$

$Y_{100} \sim B(100,0.28)$

$P(Y_{10} \leq 1) = 0.183$ `pbinom(1, 10, 0.28)`

$P(Y_{100} \leq 10)$ = 1.017339e-05, o 0.00001 `pbinom(10, 100, 0.28)`

Recuerde que la notación científica centra la atención en las cifras significativas (en este caso, 1.017339) y luego informa del cuantos ceros tendrá la división (en este caso, 5). Es decir, 1.017339 eventos cada 10^5 casos = cada 100000 = cada cien mil. 1.017339e-05 ≈ 0.00001 = 1 cada cien mil. De forma breve, este 5 marca cuantas posiciones debe moverse el símbolo decimal.

3.9. El percentil 90 será 34. Como en el 92.4% de los trimestres, el número de fracasos será como mucho de 34, podemos garantizar con una confianza del 90% (de hecho, algo superior) que el número de fracaso será 34 o menos.

3.10. Percentil 0.95 de $Y_{10} \sim B(10,0.28)$:

```
> qbinom(0.95,10,0.28)

[1] 5
```

En el 95% de las semanas, el número de fracasos será 5 como mucho.

Cuantil 0.95 de $Y_{100} \sim B(100,0.28)$:

```
> qbinom(0.95,100,0.28)

[1] 35
```

En el 95% de los trimestres, el número de fracasos será 35 como mucho.

3.11. La probabilidad de EA grave, es π=0.1 y la muestra de n=10 casos. Por lo tanto la variable X ~ B(10,0.1). R calcula la probabilidad de 0 eventos en un total de 10 casos:

```
> dbinom(0,10,0.1)

[1] 0.3486784
```

La probabilidad de no obtener ningún evento de 10 posibles con una probabilidad del 10% es del 35%.

Con la n=100, X ~ B(100,0.1), haciendo los cálculos con R obtenemos:

```
> dbinom(0,100,0.1)
[1] 2.65614e-05
```

La probabilidad de no obtener ningún caso de 100 posibles con una probabilidad de "éxito" del 10% es prácticamente 0: 0.0000265614.

Asumiendo que la probabilidad de evento sea del 10%, observar 0 eventos en 10 casos es bastante probable (35%), pero observar 0 de 100 es casi imposible (aproximadamente 3 por 100000).

3.12. Recordemos que P(X=0) con X ~ B(10,0.1):

```
> dbinom(0,10,0.1)
[1] 0.3486784
```

P(X=0) con X ~ B(100,0.01):

```
> dbinom(0,100,0.01)
[1] 0.3660323
```

P(X=0) con X ~ B(1000,0.001):

```
> dbinom(0,1000,0.001)
[1] 0.3676954
```

P(X=0) con X ~ B(1000000,0.000001):

```
> dbinom(0,1000000,0.000001)
[1] 0.3678793
```

La probabilidad de observar 0 eventos es muy parecida. De hecho, a medida que aumenta n y disminuye π, las diferencias tienden a hacerse más pequeñas y las probabilidades sucesivas más similares.

$\pi \cdot n = E(X_{10}) = E(X_{100}) = E(X_{1000}) = E(X_{1000000}) = 1$

En todos los casos en que el producto π·n es el mismo, siendo π pequeña y n grande, la esperanza es el mismo valor. Y antes vimos que la probabilidad de observar 0 eventos es muy parecida.

3.13. Se parecen mucho.

De hecho, para un mismo valor del producto π·n = E(X), se parece más cuanto más pequeña es π y mayor es n.

3.14. X ~ P(λ=1). Utilizando R obtenemos l P(X=0), P(X=1), P(X=2), P(X=3) y P(X=4). Estas probabilidades las podemos interpretar también como la frecuencia ("teórica, que cabe esperar") de días en los que se observarán ese número de eventos. Luego la buena o la mala suerte hará que oscilen alrededor de ese valor esperado. [Nota: esa suerte, esa influencia del azar, se puede cuantificar. Por ejemplo, mediante una simulación informática; o modelando, por ejemplo, la probabilidad de un valor concreto (sea 2) frente al resto (diferente de 2) como una nueva binomial.]

```
> dpois(0,1)
[1] 0.3678794
```

La probabilidad de que en un día no haya ningún caso es de 36.8%: cabe esperar que aproximadamente 1 de cada 3 días no haya trabajo.

```
> dpois(1,1)
[1] 0.3678794
```

También cabe esperar que aproximadamente 1 de cada 3 días haya 1 caso de esta urgencia.

```
> dpois(2,1)
[1] 0.1839397
```

En cambio, cabe esperar que aproximadamente 1 de cada 6 días haya 2 casos.

```
> dpois(3,1)
[1] 0.06131324
> dpois(4,1)
[1] 0.01532831
```

3.15. Como en el ejercicio anterior, $X \sim P(\lambda=1)$.

$P(X \leq 3)$:

```
> ppois(3,1)
[1] 0.9810118
```

$P(X \leq 4)$:

```
> ppois(4,1)
[1] 0.9963402
```

Las garantías de cubrir necesidades un día concreto con las dimensiones del servicio serán del 99.6%.

3.16. Para $X \sim P(\lambda=1)$, $V(X) = \lambda = 1 = \sigma^2$. Entonces $\sigma = 1$.

Para $X \sim P(\lambda=4)$, $V(X) = \lambda = 4 = \sigma^2$. Entonces $\sigma = 2$.

Tiene sentido ya que a mayor número de casos por unidad de tiempo, mayor rango de valores puede tomar la variable y por lo tanto hay más dispersión. Por otro lado, note que de forma relativa, la dispersión es menor: una $\sigma = 1$ para una $\mu=1$ es 'relativamente' mayor que una $\sigma = 2$ para una $\mu=4$; y ésta mayor que una $\sigma = 3$ para una $\mu=9$.

3.17.

```
> qpois(0.99,1)
[1] 4
```

Se tiene que estar preparado para recibir hasta 4 casos.

3.18. A) Bueno, la crítica más importante en los 3 ejemplos es la independencia. Si alguien ha tenido un accidente o una infección, quizás aumente la probabilidad de que otros también la tengan.

B) Si la tasa la hemos estimado por un buen proceso, sería correcta, lo que no sería correcto serían los valores observados.

C) Convendría comprobar empíricamente si aumenta o no aumenta. Una posibilidad sería comparar las frecuencias observadas, empíricas, a lo largo de cierto periodo de tiempo con las predichas por el modelo de Poisson.

3.19. Si se trata de una moneda no trucada la probabilidad de éxito, definido como obtener cara, es de 0,5. Por lo tanto la variable X ~ G(0.5).

P (X=0):

```
> dgeom(0,0.5)
[1] 0.5
```

P(X=1):

```
> dgeom(1,0.5)
[1] 0.25
```

La probabilidad de que salga cara a la primera es del 50%, mientras que la de tener que realizar justo dos lanzamientos hasta que salga cara es del 25%.

> pgeom (4,0.5)

[1] 0.96875

La probabilidad de obtener una cara al quinto lanzamiento o antes es muy alta, casi del 97%.

3.20. De los 10 artículos enviados, 9 previos deben ser rechazados y el décimo aceptado; como la probabilidad de aceptar un artículo es de 0.15, por tanto X ~ G(0.15); y la P(X=9)=> `dgeom(9,0.15)= 0.03474254.`

Si queremos que sea aceptado en el décimo o antes, el número de fracasos ha de ser 9 o menos: `pgeom(9,0.15)= 0.8031256`

3.21. En el primer caso, número de días que resiste sin infectarse, X_1 ~ G(0.05), y las probabilidades pedidas son

$P(X_1 \geq 10)$ = 1-pgeom (9,0.05) = 0.5987369 y

$P(X_1 \geq 20)$ = 1-pgeom (19,0.05) = 0.3584859

Si la probabilidad de infección sube hasta el 20%, los respectivos resultados serían

$P(X_1 \geq 10)$ = 1-pgeom (9,0.2) = 0.1073742 y

$P(X_1 \geq 10)$ = 1-pgeom (19,0.2) = 0.01152922.

Estos resultados son coherentes con la intuición: a más días que pasan, más aumenta la probabilidad de que un paciente adquiera una infección nosocomial. Esta es una razón importante para no alargar la estancia más de lo necesario.

3.22. La probabilidad de observar 6 caras de 10 lanzamientos si π=0.6 es

```
> dbinom(6,10,0.6)
[1] 0.2508227
```

Y la probabilidad de observar 6 caras de 10 lanzamientos si π=0.5 es

```
> dbinom(6,10,0.5)

[1] 0.2050781
```

Así, la verosimilitud de observar 6 caras en 10 lanzamientos es mayor para un valor del parámetro π=0.6 que π=0.5. Así, π=0.6 es más verosímil que π=0.5.