USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
FACULTADE DE ENFERMARÍA

Estadística y metodología de la investigación
Curso 2012-2013
Pedro Faraldo, Beatriz Pateiro

Tema 3. Variables aleatorias. Inferencia estadística

1 Introducción

Uno de los objetivos claves de la estadística es inferir o extraer conclusiones con respecto a la población basándose en la información contenida en una muestra. Para poder estudiar las características de la población es necesario dotar a la variable de interés de un modelo probabilístico de distribución que nos permita explicar su comportamiento aleatorio.

Ejemplo 1: En una ciudad de la costa gallega, los jóvenes de entre 14 y 23 años se reúnen cada noche de sábado para organizar un botellón. La ingesta de alcohol hace que un 40 % de ellos sufran una intoxicación etílica leve. Además, hay 5 de cada mil que sufren una intoxicación etílica grave (coma etílico), ya que superan los 3 g/l en sangre. A mayores de los problemas ocasionados por la intoxicación etílica, otra de las graves consecuencias son los accidentes de tráfico. La Dirección General de Tráfico estipula como tasa máxima de alochol $0.5 g/l$ en sangre (equiv. 0.25 mg/l en aire expirado).

2 Variables aleatorias

En este tema se introducen algunos resultados básicos sobre variables aleatorias y se describen dos de las familias de distribuciones discretas y continuas más relevantes: la distribución Binomial, la distribución de Poisson y la distribución Normal. Es común que los resultados posibles (espacio muestral Ω) de un experimento aleatorio no sean valores numéricos. Para el cálculo de probabilidades asociadas con un experimento resulta más sencillo utilizar valores numéricos en lugar de trabajar directamente con los elementos de un espacio muestral. De este

modo, el concepto de variable aleatoria surge ante la necesidad de representar numéricamente los resultados de un determinado experimento aleatorio.
De manera simplificada, podríamos decir que una variable aleatoria es una correspondencia que asocia a cada elemento del espacio muestral de un experimento un número. Dependiendo de las posibles asignaciones numéricas a la variable, distinguiremos entre variables aleatorias discretas y continuas.

2.1 Variable aleatoria discreta.

Si X es una variable aleatoria (v.a.) sobre un espacio muestral Ω, y sólo toma valores en un conjunto finito (o infinito numerable) entonces diremos que X es una variable aleatoria discreta. Si $x_1, \ldots, x_n$ son los posibles valores que toma una v.a. discreta, al conjunto de probabilidades $p_1, \ldots, p_n$ tales que:

$$\mathbb{P}(X = x_i) = p_i, i = 1, \ldots, n \quad \text{con} \quad \sum_{i=1}^{n} p_i = 1$$

se le denomina **función de masa de probabilidad**. El comportamiento de una variable aleatoria también se puede describir a través de la **función de distribución**. La función de distribución de una v.a. X es una función que a cada valor real x le asocia la probabilidad de que la variable tome valores menores o iguales a dicho número:

$$F(x) = \mathbb{P}(X \leq x).$$

Es decir, la función de distribución de una v.a. X en un punto x nos da la probabilidad acumulada hasta este valor. Esta función toma valores entre 0 y 1, y es no decreciente.

2.2 Medidas características.

Sea X una v.a. discreta, con valores $x_1, \ldots, x_n$ y masa de probabilidad $p_1, \ldots, p_n$. Podemos obtener las siguientes medidas características:

1. La media o esperanza matemática:

$$\mathbb{E}(X) = \mu = \sum_{i=1}^{n} x_i p_i.$$

Propiedades de la media:

a) $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b, \quad a, b \in \mathbb{R}$.
b) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.
c) Si g es una función: $\mathbb{E}(g(X)) = g(\mathbb{E}(X)) = \sum_{i=1}^{n} g(x_i)p_i$.
d) $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, si X, Y son independientes (*Para la definición de independencia de variables, véase Crujeiras y Faraldo (2010)*).

2. La varianza y la desviación típica:

$$\sigma^2 = \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \quad \text{y} \quad \sigma = +\sqrt{\mathbb{E}[(X - \mathbb{E}(X))^2]}.$$

Propiedades de la varianza:

a) $\text{Var}(aX + b) = a^2\text{Var}(X)$.
b) $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mu^2$.
c) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$, si X e Y son independientes.

3. El coeficiente de variación (*Se define de igual manera para v.a. continuas.*): $CV = \dfrac{\sigma}{\mu}$.
4. La mediana es el valor Me que divide la distribución en dos partes iguales (*Se define de igual manera para v.a. continuas.*). Es decir: $F(Me) = \dfrac{1}{2}$.
5. La moda es el valor donde la función de masa de probabilidad alcanza su máximo.

2.3 Distribución Binomial

Un experimento de Bernoulli es aquel que sólo presenta dos posibles resultados (por ejemplo, éxito o fracaso, válido o defectuoso, 0 o 1, sano o enfermo, etc). En general, llamaremos éxito (E) a la ocurrencia del suceso que nos interesa estudiar, y fracaso (F) a la no ocurrencia. Además, la probabilidad de obtener un éxito se mantiene constante, y la denotaremos por p. La v.a. X que registra el resultado de una prueba de este tipo se dice que tiene una distribución de Bernoulli de parámetro p: $X \sim \text{Ber}(p)$:

$$X = \begin{cases} 1 & \text{si ocurre E,} \\ 0 & \text{si ocurre F,} \end{cases}$$

o también:

$$X = \begin{cases} 1 & \text{con probabilidad } p, \\ 0 & \text{con probabilidad } q = 1 - p. \end{cases}$$

La función de masa de probabilidad de $X \sim \text{Ber}(p)$ es:

x_i	1	0
p_i	p	$q = 1 - p$

En general, podemos escribir la función de masa de probabilidad como:

$$\mathbb{P}(X = x) = p^x(1-p)^{1-x} = p^x q^{1-x}.$$

La esperanza y la varianza son:

$$\mathbb{E}(X) = p, \quad \text{Var}(X) = pq.$$

Si repetimos el experimento de Bernoulli n veces y consideramos la variable que cuenta el número de éxitos esta seguirá una **distribución Binomial** de parámetros n y p ($X \sim \text{Bi}(n, p)$):

$$X = \text{nº de éxitos en } n \text{ pruebas de Bernoulli} \Rightarrow X \sim \text{Bi}(n, p)$$

Esta variable, puede tomar valores $\{0, 1, 2, \ldots, n-1, n\}$. La función de masa de probabilidad es:

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

La esperanza y la varianza son:

$$\mathbb{E}(X) = np, \quad \text{Var}(X) = npq = np(1-p).$$

Propiedades de la Binomial:

1. La distribución de Bernoulli es una Binomial con $n = 1$: $\text{Ber}(p) = \text{Bi}(1, p)$.
2. La masa de probabilidad está tabulada, al igual que la función de distribución.

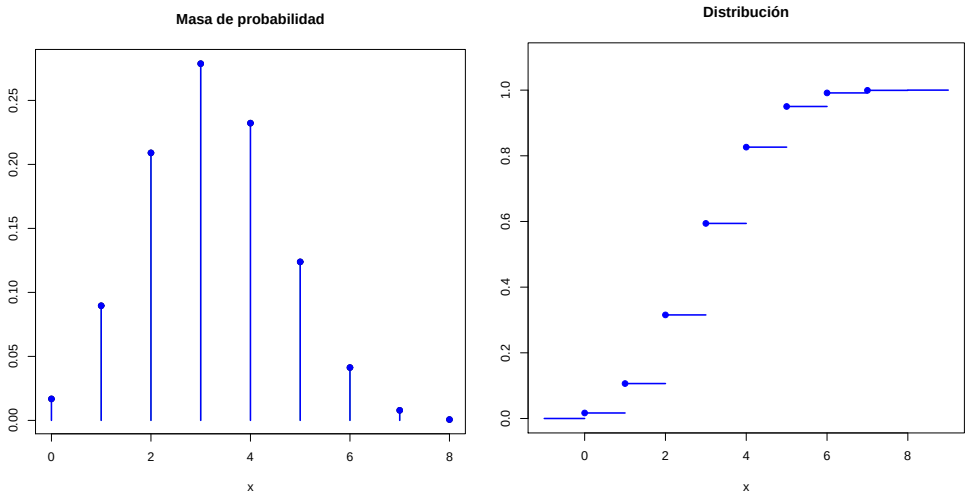

Figura 1: Masa de probabilidad y función de distribución de una Bi(8, 0.4).

3. Si $X \sim \text{Bi}(n_1, p)$, $Y \sim \text{Bi}(n_2, p) \Rightarrow X + Y \sim \text{Bi}(n_1 + n_2, p)$ (con X e Y independientes).
4. Si $X \sim \text{Bi}(n, p)$ entonces $X = \sum_{i=1}^{n} X_i$, donde $X_i \sim \text{Ber}(p)$.

■ Volviendo al ejemplo sobre un grupo de 6 amigos del ejemplo, calcula:

a) La probabilidad de que 3 sufran intoxicación etílica leve.

b) La probabilidad de que al menos 2 sufran intoxicación etílica leve.

c) Número esperado de amigos que sufrirán intoxicación etílica leve.

En este caso, encontrar a alguien que sufra intoxicación etílica leve es lo que hemos denominado éxito. Por tanto, $p = 0.4$ (el 40 % sufren intoxicación etílica leve). Si definimos la variable:

$$X = \{\text{nº de amigos, en el grupo de 6, que sufren intoxicación etílica leve}\}$$

esta sigue una distribución $X \sim \text{Bi}(6, 0.4)$. Por tanto, para resolver el apartado a) tendríamos que calcular:

$$\mathbb{P}(X = 3) = \binom{6}{3} 0.4^3 0.6^{6-3} = 0.276.$$

En el apartado b), necesitamos calcular $\mathbb{P}(X \geq 2)$, que se podría hacer como la suma de $\mathbb{P}(X = 2)$, $\mathbb{P}(X = 3)$, ... hasta $\mathbb{P}(X = 6)$, o bien:

$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X < 2) = 1 - [\mathbb{P}(X = 0) + \mathbb{P}(X = 1)] = 1 - 0.23 = 0.77.$$

Finalmente, el número esperado de amigos que sufrirán intoxicación etílica leve será al media de la variable X:

$$\mathbb{E}(X) = n \cdot p = 6 \cdot 0.4 = 2.4.$$

Observa que, aunque la variable sea discreta con $Sop(X) = \{0, 1, \ldots, 6\}$, la esperanza no tiene porqué ser un valor del soporte, pero debe estar entre el máximo y el mínimo de los posibles valores.

2.4 Distribución de Poisson.

Un proceso de Poisson es un experimento aleatorio que consiste en observar la aparición de sucesos en un soporte continuo, por ejemplo, en el tiempo. Este proceso ha de ser estable: es decir, el número medio de sucesos por unidad de tiempo (λ) se mantiene constante. Además, los sucesos han de ser independientes. Si consideramos la variable:

$$X = \text{nº de sucesos en un intervalo} \Rightarrow X \sim Pois(\lambda).$$

Esta variable toma valores $\{0, 1, 2, \ldots\}$. La masa de probabilidad es:

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

La media y la varianza de una v.a. de Poisson son:

$$\mathbb{E}(X) = \text{Var}(X) = \lambda.$$

■ Siguiendo con el ejemplo, en los servicios de urgencias del hospital más cercano se registra una llegada media de 2 personas cada 10 minutos, por intoxicación etílica. Calcula:

a) El número esperado de personas que llegarán en los próximos 20 minutos.

b) Probabilidad de que en los próximos 20 minutos lleguen 5 personas.

c) Probabilidad de que en los próximos 20 minutos lleguen, al menos, 3 personas.

Si definimos la variable:

$$X = \{\text{nº de personas que llegan, por intoxicación etílica, cada 20 minutos}\}$$

esta variable tendrá una distribución $X \sim Pois(4)$ (si en 10 minutos se registra una llegada media de 2 personas, en 20 minutos se tendrá una media de 4 personas). Por tanto, el número esperado $\mathbb{E}(X) = 4$. Para resolver el apartado b), debemos calcular:

$$\mathbb{P}(X = 5) = \frac{e^{-4}4^5}{5!} = 0.156.$$

Finalmente, se pide $\mathbb{P}(X \geq 3)$. Para resolver esto, debemos tener en cuenta que:

$$\begin{aligned}
\mathbb{P}(X \geq 3) &= 1 - \mathbb{P}(X < 3) \\
&= 1 - \mathbb{P}(X \leq 2) \\
&= 1 - [\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2)] \\
&= 0.762.
\end{aligned}$$

2.5 Variable aleatoria continua.

Una v.a. continua es aquella que toma valores en un intervalo (o varios intervalos) de la recta real. La función de distribución de una v.a. continua se define de igual manera a la de una v.a. discreta, es decir, la función de distribución de una v.a. X es una función que a cada valor real x le asocia la probabilidad de que la variable tome valores menores o iguales a dicho número, al igual que para variables discretas $F(x) = \mathbb{P}(X \leq x)$. Como

generalización al caso continuo de la función de masa de probabilidad se tiene la **función de densidad**. Dada una v.a. continua X, la función de densidad se define como:

$$f(x) = \lim_{h \to 0} \frac{\mathbb{P}(x - h < X < x + h)}{2h}$$

Si la función de distribución de X es F, entonces tenemos la siguiente relación entre densidad y distribución:

$$f(x) = F'(x), \quad \text{o también } F(x) = \int_{-\infty}^{x} f(t)dt.$$

Propiedades de la función de densidad:

1. Dado que la función de densidad es la derivada de la distribución, y ésta es una función no decreciente: $f(x) \geq 0$, $-\infty < x < \infty$.
2. El área bajo la curva de densidad integra 1.
3. La probabilidad de que X tome un valor concreto es nula.
4. La probabilidad de un intervalo (a, b) es el área bajo la curva $f(x)$ entre las rectas $x = a$ y $x = b$. Es decir:
$$\mathbb{P}(a < X < b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$$
Por tanto, la probabilidad de (a, b) es la misma que la de $[a, b]$, $[a, b)$ o $(a, b]$.

2.6 Medidas características.

Si X es una v.a. continua, con función de densidad f se definen:

1. La media o esperanza matemática (*Las propiedades son las mismas que las estudiadas en el caso de v.a. discretas.*):
$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$$
2. La varianza y la desviación típica:
$$\sigma^2 = \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$
$$\sigma = +\sqrt{\int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx}$$
La varianza admite un cálculo más sencillo de la siguiente forma:
$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

2.7 Distribución Normal

Una v.a. X tiene distribución Normal con media μ y varianza σ^2 si su densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Si tenemos $\mu = 0$ y $\sigma^2 = 1$, entonces $X \sim N(0, 1)$ y su densidad será:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

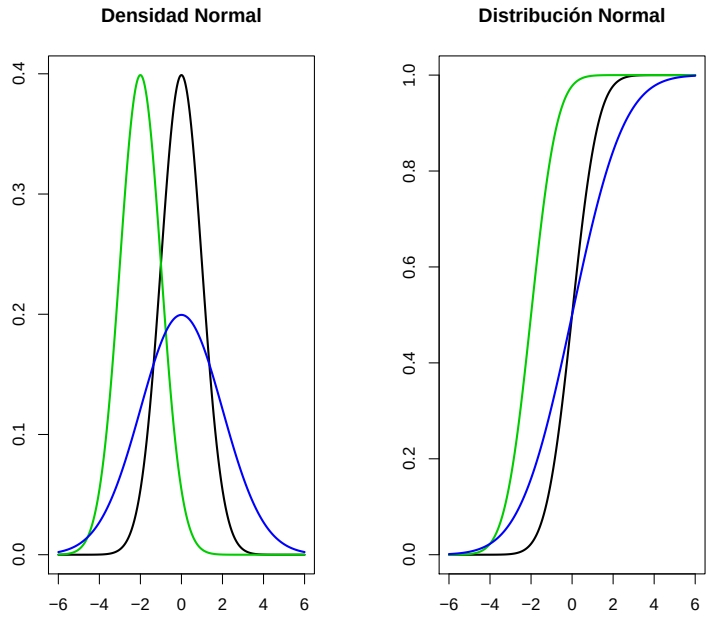

Figura 2: Densidad y distribución Normal. Negro: $N(0, 1)$. Verde: $N(-2, 1)$. Azul: $N(0, 4)$.

Como podemos ver en la Figura 2, el cambiar los valores de los parámetros de localización (media) y escala (desviación típica), tiene distinto efecto sobre la forma de la densidad y por tanto, de la distribución.

Tomando como referencia la Normal estándar ($N(0, 1)$, distribución Normal con media $\mu = 0$ y varianza $\sigma^2 = 1$), si cambiamos la media, lo que hacemos es trasladar la gráfica, hacia la derecha si la media es positiva, y hacia la izquierda si es negativa.
Cuando modificamos la varianza, si la aumentamos lo que estamos haciendo es incrementar la dispersión, con lo que la curva se *achata*, incrementando la probabilidad de los valores más altos y más bajos con respecto al modelo estándar. Si la reducimos, lo que veríamos es que se concentra más alrededor de la media. Al incremetar la varianza, la densidad se vuelve mesocúrtica (curtosis negativa), mientras que al disminuir la varianza, lo que se obtiene es una curva leptocúrtica (curtosis positiva). La densidad de una $N(0, 1)$ tiene curtosis nula (platocúrtica).

La función de distribución de la Normal estándar, que denotaremos por $\Phi(z)$, está tabulada. En la distribución $N(0, 1)$ será de utilidad que identifiquemos en qué intervalos se encuentran el 90 %, 95 % y 99 % de los valores. En la Figura 3 se muestran los tres intervalos más usuales para los posibles valores de una $N(0, 1)$. Estos intervalos serán de utilidad tanto en la estimación por intervalos de confianza como para los contrastes de hipótesis.

Cuando la distribución de la variable es Normal, pero no estándar, para poder calcular probabilidades a partir de las tablas necesitamos tipificar. Es decir, si $X \sim N(\mu, \sigma^2)$, la transformada:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

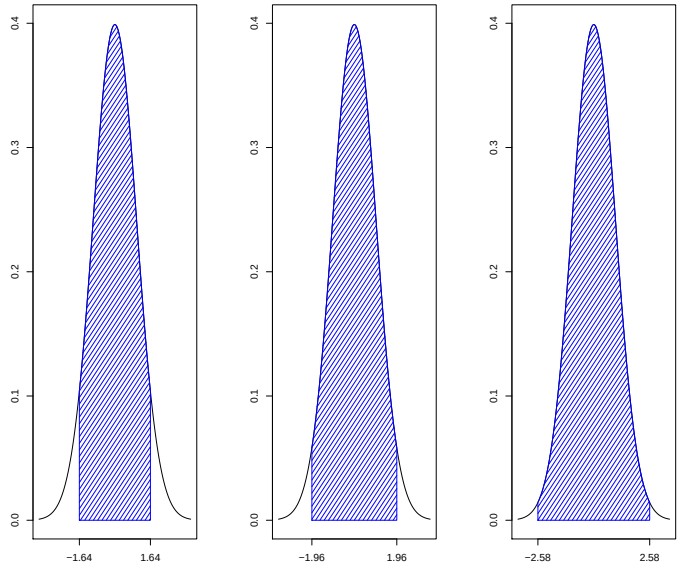

Figura 3: Intervalos en una $N(0,1)$. El 90 % de los valores están en $(-1.64, 1.64)$. El 95 % de los valores están en $(-1.96, 1.96)$. El 99 % de los valores están en $(-2.58, 2.58)$.

■ Veíamos en el ejemplo que al cabo de tres horas, la concentración media de alcohol en sangre de los jóvenes es de 0.45 g/l, con una desviación típica de 0.4 g/l. Si esta concentración se distribuye según una Normal, calcula:

a) La probabilidad de que un individuo elegido al azar no supere los 0.6 g/l.

b) La probabilidad de que un individuo que tiene más de 0.2 g/l pueda conducir.

Si denotamos por $X = \{\text{concentración de alcohol en sangre, al cabo de 3 horas}\}$, esta variable tiene una distribución $X \sim N(\mu = 0.45, \sigma^2 = 0.4 \cdot 0.4 = 0.16)$. Para calcular estas probabilidades, tendremos que tipificar:

$$\mathbb{P}(X \leq 0.6) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{0.6-0.45}{0.4}\right) = \mathbb{P}(Z \leq 0.375) = 0.646.$$

Para el segundo apartado, debemos tener en cuenta que un individuo puede conducir si su tasa de alcohol es menor de 0.5 g/l. Por tanto, calcularemos:

$$\mathbb{P}(0.2 < X < 0.5) = \mathbb{P}\left(\frac{0.2-0.45}{0.4} < Z < \frac{0.5-0.45}{0.4}\right) =$$
$$\mathbb{P}(-0.625 < Z < 0.125) = \mathbb{P}(Z \leq 0.125) - \mathbb{P}(Z \leq -0.625) =$$
$$0.549 - 0.266 = 0.263.$$

Ten en cuenta que, dado que la Normal es continua, la probabilidad puntual es nula, y podemos utilizar $<$ o $\leq$ ($>$ o $\geq$) indistintamente.

En algunos casos lo que nos interesa no es calcular la probabilidad de unos ciertos valores sino, dada una probabilidad, ver cuál es el punto de corte que deja esa proporción de valores de la variable por encima o por

debajo. Para ello, se define la función de distribución inversa o la función cuantil del siguiente modo. Dada una probabilidad p_0, la función cuantil q nos devuelve el punto x_0 tal que:

$$q(p_0) = x_0, \quad \text{si} \quad \mathbb{P}(X \leq x_0) = F(x_0) = p_0.$$

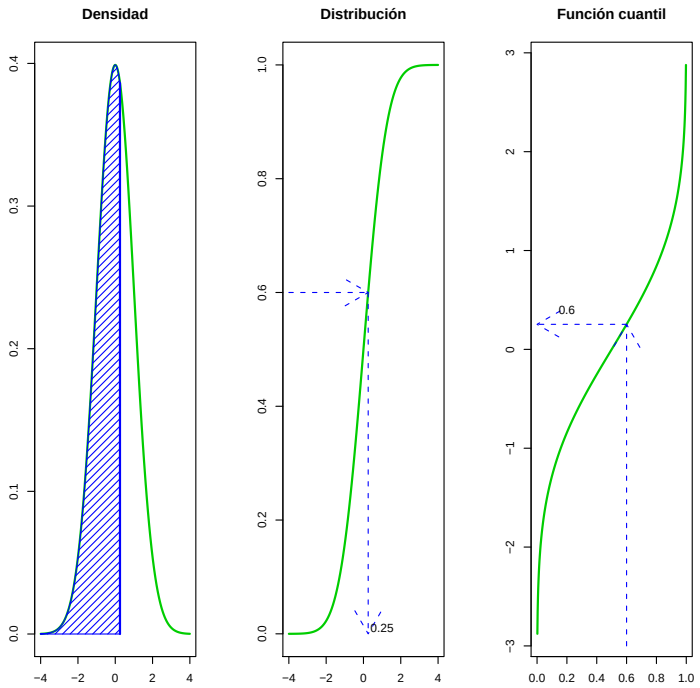

Figura 4: Densidad, distribución y función cuantil (distribución inversa) de una $N(0,1)$. Ejemplo para el sexto decil (cuantil 0.6). $\mathbb{P}(Z \leq z_0) = 0.6 \Rightarrow z_0 = 0.2533$.

La distribución Normal también satisface la siguiente propiedad de aditividad. Si $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ son independientes, entonces $(X + Y) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

- Una aplicación de los cuantiles en nuestro ejemplo sería la siguiente:

  a) La concentración máxima para el 20 % de los individuos con menos alcohol en sangre.

  b) La concentración mínima para el 15 % de los individuos con más alcohol en sangre.

La variable de interés, concentración de alcohol en sangre, es la misma que en el desarrollo anterior: $X \sim N(0.45, 0.16)$. Para el apartado a), debemos obtener el punto x_0 que verifica:

$$\mathbb{P}(X \leq x_0) = 0.2.$$

Al igual que para el cálculo de probabilidades, tendremos que tipificar:

$$\mathbb{P}(X \leq x_0) = 0.2 \Leftrightarrow \mathbb{P}\left(Z \leq \frac{x_0 - 0.45}{0.4}\right) = 0.2.$$

El cuantil 0.2 de una $N(0,1)$ es: $q(0.2) = -0.84$. Por tanto:

$$\frac{x_0 - 0.45}{0.4} = -0.84 \Leftrightarrow x_0 = 0.45 - 0.84 \cdot 0.4 = 0.114 \ g/l.$$

Por tanto, el 20 % de los individuos tienen una tasa de alcohol en sangre inferior a 0.114 g/l. Para el segundo apartado, procederíamos de forma similar, pero teniendo en cuenta:

$$\mathbb{P}(X \geq x_0) = 0.15 \Leftrightarrow \mathbb{P}(X \leq x_0) = 0.85.$$

En algunos casos, tendremos que utilizar ambas la distribución Normal y la Binomial para poder responder a las cuestiones que se nos planteen. Veamos cómo se resolvería el siguiente ejemplo:

■ Con los datos obtenidos anteriormente, para un grupo de 6 amigos mayores de 18 años, calcula la probabilidad de que sólo uno pueda conducir.

Si consideramos $X = \{$nº de amigos, en el grupo de 6, que pueden conducir$\}$, esta variable tendrá una distribución $X \sim \text{Bi}(6, p)$, donde p es la probabilidad de que un individuo pueda conducir, o equivalentemente, que tenga una tasa de alcohol en sangre inferior a 5 g/l. ¿Cómo calculamos p?

Consideramos $Y = \{$tasa de alcohol en sangre$\}$, con distribución $Y \sim N(0.45, 0.16)$. Entonces,

$$p = \mathbb{P}(Y \leq 0.5) = \mathbb{P}\left(Z \leq \frac{0.5 - 0.45}{0.5}\right) = \mathbb{P}(Z \leq 0.125) = 0.55.$$

Así, $X \sim \text{Bi}(6, 0.55)$, y la probabilidad de que sólo uno pueda conducir se tendría como:

$$\mathbb{P}(X = 1) = \binom{6}{1} 0.55^1 0.45^5 = 0.06.$$

---

3 Introducción a la inferencia estadística

Una vez introducidos los modelos de probabilidad que describen el comportamiento de las poblaciones de interés, veremos algunos de los conceptos básicos de la inferencia estadística, que tiene como objetivo extraer conclusiones sobre la población basándose en la información contenida en una muestra. Entre los problemas que se pretenden resolver con la inferencia estadística se distinguen dos tipos: la estimación, tanto puntual como por intervalos y los contrastes de hipótesis, que se abordarán en los siguientes temas.

- **Población:** conjunto homogéneo de individuos sobre los que se estudian características observables con el objetivo de extraer alguna conclusión.
- **Parámetro:** característica de la población, por ejemplo, la media, la varianza,...
- **Estadístico:** cualquier función de la muestra. Por ejemplo, la media o la varianza muestrales son estadísticos. Los estadísticos los denotaremos por $T(X_1, \ldots, X_n)$.

- **Estimadores:** son estadísticos independientes de los parámetros de la población, y que se utilizan para aproximarlos. Si θ es el parámetro de interés, el estimador se denotará por $\hat{\theta}$. Por ejemplo, podemos considerar la media muestral como estimador de la media poblacional:

$$T(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} = \hat{\mu}.$$

- **Método de muestreo:** procedimiento para seleccionar una muestra. Si en una población queremos obtener una muestra de un cierto tamaño n (siendo n menor que el tamaño de la población), la manera de obtener esta muestra no es única. En la siguiente sección, describiremos distintos métodos para seleccionar muestras.

3.1 Tipos de muestreo

En esta sección describiremos brevemente cuatro métodos de muestreo clásicos: muestreo aleatorio simple, muestreo sistemático, muestreo estratificado y muestreo por conglomerados.
En los desarrollos posteriores que realicemos, consideraremos que la muestra de la que disponemos se ha obtenido mediante muestreo aleatorio simple.

- **Muestreo aleatorio simple:** cada muestra posible de tamaño n tiene la misma probabilidad de ser seleccionada, y cada individuo de la población tiene la misma probabilidad de caer en la muestra.

- **Muestreo sistemático:** se utiliza cuando los individuos están ordenados en listas. Si tenemos una población de N individuos y queremos extraer una muestra de tamaño n, debemos calcular k (parte entera de N/n) y elegir un valor l en $\{1, 2, \ldots, k\}$. Los elementos de la muestra se seleccionan como aquellos en las posiciones $\{l, k + l, 2k + l, \ldots, (n - 1)k + l\}$.

- **Muestreo estratificado:** cuando en la población existen grupos o clases homogéneos con respecto a la característica a estudiar (estratos), los individuos de la muestra se seleccionan con una cierta *afijación* en cada estrato. Es decir, para seleccionar una muestra de tamaño n, si tenemos K estratos en la población, elegiremos en cada uno $n_1, \ldots, n_K$ individuos, de tal modo que $\sum_{k=1}^{K} n_k = n$. En cada estrato, los n_k se suelen seleccionar por muestreo aleatorio o por muestreo sistemático.

  - *Afijación simple:* $n_1 = n_2 = \ldots = n_K$.
  - *Afijación proporcional:* cada n_k es proporcional al tamaño del estrato en la población.

- **Muestreo por conglomerados:** los conglomerados son subgrupos de la población homogéneos entre sí. Se eligen aleatoriamente algunos conglomerados y en cada uno de ellos se estudia a toda la población (o se selecciona una muestra, en cuyo caso se denomina muestreo bietápico).

4 Teorema Central del Límite

El Teorema Central del Límite es uno de los principales resultados en la teoría de la probabilidad. En su enunciado más simple, el Teorema Central del Límite establece que la suma de un número grande de observaciones independientes de la misma distribución se aproxima a una distribución Normal.

Teorema 1 (Teorema Central del Límite)**.** *Sean $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2. Si $n \to \infty$ (n suficientemente grande) entonces:*

$$S_n = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2).$$

Observa que para obtener este resultado, no es necesario que las variables X_i tengan distribución Normal, si no que el resultado es válido para cualquier $X \sim F$. Es suficiente con que tengan la misma distribución.

5 Aproximaciones entre distribuciones

Cuando el número de pruebas del experimento de Bernoulli n es grande, podemos hacer aproximaciones de la distribución Binomial que facilitarán el cálculo de probabilidades, bien a la distribución de Poisson o bien a la Normal.
En la Figura 5 podemos ver gráficamente la validez de la aproximación de la Binomial a la Normal. En la parte izquierda tenemos la masa de probabilidad de una Bi(50, 0.3). Como el número de pruebas n es suficientemente grande (se suele fijar el criterio $n > 30$), podríamos aproximar la Binomial por una Normal $N(\mu, \sigma^2)$ con $\mu = np = 50 \cdot 0.3 = 15$ y $\sigma^2 = npq = 50 \cdot 0.3 \cdot 0.7 = 75$.

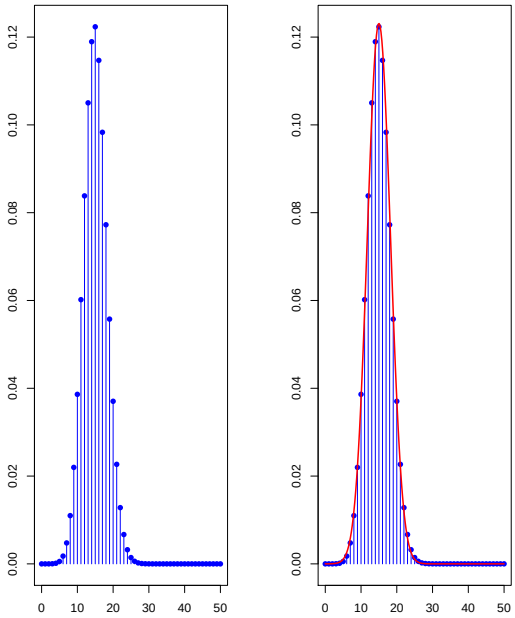

Figura 5: Masa de probabilidad de una Bi(50, 0.3) y función de densidad de una $N(15, 75)$.

Distribución	Caso	Aproximación
$Bi(n, p)$	$n \geq 30$, $p < 0.1$	$Pois(np)$
$Bi(n, p)$	$n \geq 30$, $0.1 < p < 0.9$	$N(np, npq)$
$Pois(\lambda)$	$\lambda \geq 10$	$N(\lambda, \lambda)$

Cuadro 1: Aproximaciones de la Binomial, Poisson y Normal.

Cuando en una distribución Binomial con n grande, la probabilidad de éxito es extrema ($p < 0.1$ o $p > 0.9$), se aproxima por una Poisson, con la misma media, es decir, $Pois(np)$.
En el caso de la distribución de Poisson, cuando el parámetro λ es grande (normalmente se toma $\lambda > 10$), se puede aproximar por una Normal que conserva la media y la varianza. Es decir, una variable $Pois(\lambda)$ se aproximaría por una $N(\lambda, \lambda)$, para λ suficientemente grande.

Corrección de Yates

Al aproximar la distribución Binomial o la Poisson por la Normal estamos conservando la media y la varianza. Sin embargo, al calcular probabilidades utilizando la aproximación a la Normal debemos tener en cuenta que tanto la Binomial como la Poisson son discretas, mientras que la Normal es continua.
Para una variable $X \sim \text{Bi}(50, 0.3)$ podemos calcular la probabilidad de que X sea 20, $\mathbb{P}(X = 20)$ y podemos ver en la Figura 5 que es positiva. Sin embargo, utilizando la aproximación $X \sim N(15, 75)$, al ser continua, la probabilidad $\mathbb{P}(X = 20)$ será nula. Para solucionar este problema, utilizaremos la corrección de Yates. La probabilidad de que X sea igual a 20 es la misma que:

$$\mathbb{P}(X = 20) = \mathbb{P}(19.5 < X < 20.5),$$

y sobre esta segunda expresión podemos utilizar la aproximación Normal, obteniendo un resultado no nulo. Esta corrección también se debe emplear al aproximar la Poisson por la Normal.

■ *Aproximación de la Binomial por la Normal.* Sobre el ejemplo, en un grupo de 200 personas mayores de 18 años, calcula:

1. La probabilidad de que 100 puedan conducir.
2. La probabilidad de que el número de posibles conductores esté entre 80 y 105 (no incluidos).

Consideramos la variable

$$X = \{\text{nº de personas, en el grupo de 200, que pueden conducir}\}.$$

Esta variable será $X \sim \text{Bi}(200, p)$, donde p es la probabilidad de que una persona pueda conducir (es decir, que su tasa de alcohol en sangre sea inferior a 0.5 g/l). En el Capítulo 3, vimos que $p = 0.55$, teniendo en cuenta que la tasa de alcohol en sangre se distribuía según una $N(0.45, .16)$. Por tanto, $X \sim \text{Bi}(200, 0.55)$ (con $\mathbb{E}(X) = 110$, $\text{Var}(X) = 49.5$) y debemos calcular $\mathbb{P}(X = 110)$.

Dado que $n = 200$ y $p \in (0.1, 0.9)$, aproximaremos la distribución por una $N(110, 49.5)$. Como estamos pasando de una variable discreta a una continua, debemos hacer la corrección de Yates:

$$\begin{aligned}
\mathbb{P}(X = 110) &= \mathbb{P}(99.5 < X < 100.5) \\
&= \mathbb{P}\left(\frac{99.5 - 110}{\sqrt{49.5}} < Z < \frac{100.5 - 110}{\sqrt{49.5}}\right) \\
&= \mathbb{P}(-1.49 < Z < -1.35) \\
&= \mathbb{P}(1.35 < Z < 1.49) \\
&= 0.02.
\end{aligned}$$

Para el segundo apartado, debemos aplicar la corrección en ambos extremos del intervalo:

$$\mathbb{P}(80 < X < 105) = \mathbb{P}(80.5 \leq X \leq 104.5),$$

y aquí se aplicaría la aproximación a la $N(110, 49.5)$, obteniendo, después de tipificar:

$$\mathbb{P}(80.5 \leq X \leq 104.5) = \mathbb{P}(-4.19 < Z < -0.78) = 0.22.$$

■ *Aproximación de la Binomial por la Poisson.* Sobre el ejemplo del Capítulo 3, en un grupo de 200 personas, calcula la probabilidad de que 2 de ellas sufran intoxicación etílica grave.

En este caso, la variable:

$$X = \{\text{nº de personas, en el grupo de 200, que sufren intoxicación etílica grave}\}$$

sigue una distribución Bi(200, 0.005) (5 de cada 1000 sufren intoxicación etílica grave). Aunque n es suficientemente grande, no podremos aproximar esta distribución por una Normal, ya que la probabilidad de éxito es extrema. La aproximación se hará a una Poisson con media $np = 200 \cdot 0.005 = 1$. Es decir, $X \sim Pois(1)$. Por tanto:

$$\mathbb{P}(X = 2) = \frac{e^{-1}1^2}{2!} = 0.184.$$

Anexo. Tabla de relaciones

Estadística descriptiva	**V.A. discreta**	**V.A.continua**
Muestra $(x_1, \ldots, x_n)$	X v.a. discreta $Sop(X) = \{x_1, \ldots, x_n\}$	X v.a. continua $Sop(X) = (a, b) \subseteq \mathbb{R}$
$f_i = \dfrac{n_i}{n}$ frecuencia relativa $f_i \geq 0, \sum_{i=1}^{n} f_i = 1$	$p_1, \ldots, p_n$ masa de probabilidad $p_i \geq 0, \sum_{i=1}^{n} p_i = 1$	$f(x)$ función de densidad $f(x) \geq 0, \int_{-\infty}^{\infty} f(x)dx = 1$
$F_i = \dfrac{N_i}{n}$ frec. relativa acumulada $F_k = 1$ (k clases)	$F(x) = \mathbb{P}(X \leq x)$ distribución $F(-\infty) = 0, \quad F(+\infty) = 1$	$F(x) = \mathbb{P}(X \leq x)$ distribución $F(-\infty) = 0, \quad F(+\infty) = 1$
$\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$	$\mu = \mathbb{E}(X) = \sum_{i=1}^{n} x_i p_i$	$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$
$s^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ $s^2 = \dfrac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$	$\sigma^2 = \mathrm{Var}(X) = \sum_{i=1}^{n}(x_i - \mu)^2 p_i$ $\sigma^2 = \sum_{i=1}^{n} x_i^2 p_i - \mu^2$	$\sigma^2 = \mathrm{Var}(X) = \int_{-\infty}^{\infty}(x - \mu)^2 f(x)dx$ $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$

Cuadro 2: Tabla de relaciones: estadística descriptiva, variable aleatoria discreta y variable aleatoria continua.