

MODELOS DE PROBABILIDAD Y MUESTREO ALEATORIO

Julián de la Horra

Departamento de Matemáticas U.A.M.

1 Introducción

La Estadística Descriptiva nos ofrece una serie de herramientas muy útiles para resumir gráfica y numéricamente los datos que hemos obtenido sobre una característica o variable de interés, X , de una población. Estos resúmenes son muy interesantes, pero el objetivo de la Estadística habitualmente va más allá: pretende obtener conclusiones sobre la población a partir de los datos obtenidos en la muestra. La obtención de conclusiones será el objetivo de la Inferencia Estadística y para su desarrollo necesitaremos los modelos de probabilidad. En particular, será necesario modelizar las variables de interés, X , como variables aleatorias. En este capítulo, presentaremos el concepto de variable aleatoria (discreta y continua), y los modelos de probabilidad más utilizados en la práctica. Finalmente, introduciremos las características que debe poseer una muestra aleatoria, para que podamos obtener conclusiones razonadas sobre toda la población.

2 Variables aleatorias discretas

Una variable aleatoria discreta es una modelización de una característica X de tipo discreto.

Recordemos que una característica X es de tipo discreto cuando puede tomar una serie de valores claramente separados x_1, \dots, x_k . En una muestra concreta de tamaño n , cada uno de estos valores aparece n_1, \dots, n_k veces (frecuencias absolutas). La frecuencia relativa de cada valor es $f_i = n_i/n$.

Definición.- Una **variable aleatoria**, X , decimos que es de **tipo discreto** cuando puede tomar los valores x_1, \dots, x_k con probabilidades $P(x_1), \dots, P(x_k)$. Estas probabilidades reciben el nombre de **función de masa o función de probabilidad**.

Las probabilidades son la modelización en la población de las frecuencias relativas. Igual que las frecuencias relativas, las probabilidades son números entre 0 y 1, y la suma de las probabilidades es 1.

Ejemplo.- Consideremos la variable X = “Resultado obtenido” al lanzar un dado corriente con seis caras. Podemos enfrentarnos a esta variable de dos maneras diferentes:

1. **Datos muestrales.** Lanzamos la moneda n veces, y anotamos los resultados: obtenemos n_1 veces el número 1, ..., n_6 veces el número 6. La frecuencia relativa con la que hemos obtenido el valor i es $f_i = n_i/n$. Si lanzamos el dado muchas veces, seguramente las frecuencias relativas serán todas ellas bastante parecidas a $1/6$, si el dado está equilibrado.
2. **Modelo teórico.** Consideramos la variable X = “Resultado obtenido” como una variable aleatoria discreta que puede tomar los valores 1, ..., 6, cada uno de ellos con probabilidad $1/6$.

Cuando trabajábamos con variables discretas en Estadística Descriptiva, podíamos calcular la media muestral y la varianza muestral. Si obteníamos los valores $x_1, \dots, x_k, n_1, \dots, n_k$ veces, respectivamente, teníamos:

$$\text{Media muestral} = \bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k f_i x_i$$

$$\text{Varianza muestral} = v_x = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \sum_{i=1}^k f_i (x_i - \bar{x})^2$$

Las definiciones de media y varianza para una variable aleatoria discreta siguen la misma filosofía, sustituyendo frecuencias relativas por probabilidades.

Definiciones.- Consideramos una variable aleatoria discreta X que puede tomar los valores x_1, \dots, x_k con probabilidades $P(x_1), \dots, P(x_k)$.

La **media o esperanza** de X se define como:

$$\mu = E[X] = \sum_{i=1}^k x_i P(x_i)$$

La **varianza** de X se define como:

$$\sigma^2 = V(X) = \sum_{i=1}^k (x_i - E[X])^2 P(x_i) = \dots = \sum_{i=1}^k x_i^2 P(x_i) - (E[X])^2$$

3 Variables aleatorias continuas

Una variable aleatoria continua es una modelización de una característica X de tipo continuo.

Recordemos que una característica X es de tipo continuo cuando puede tomar cualquier valor en un intervalo de la recta real. En una muestra concreta de tamaño n , los valores obtenidos se pueden representar gráficamente, en un diagrama de tallos y hojas o en un histograma, obteniendo así el perfil de los datos.

Definición.- Una **variable aleatoria**, X , decimos que es de **tipo continuo** cuando puede tomar cualquier valor en un intervalo de la recta real con una **función de densidad** $f(x)$ que representa la idealización en la población del perfil obtenido a partir de los datos en el diagrama de tallos y hojas o en el histograma.

Las propiedades básicas de cualquier función de densidad son las siguientes:

1. $f(x) \geq 0$ (las frecuencias relativas tampoco podían ser negativas).
2. $\int_{\mathfrak{R}} f(x)dx = 1$ (las frecuencias relativas también sumaban uno).
3. $P(X \in I) = \int_I f(x)dx$ (la función de densidad sirve para calcular la probabilidad de que la variable aleatoria X tome valores en un intervalo I que nos interese).

La media y la varianza de una variable aleatoria continua se definen como en el caso discreto, sustituyendo las probabilidades por la función de densidad:

Definiciones.- Consideramos una variable aleatoria continua X con función de densidad $f(x)$.

La **media o esperanza** de X se define como:

$$\mu = E[X] = \int_{\mathfrak{R}} xf(x)dx$$

La **varianza** de X se define como:

$$\sigma^2 = V(X) = \int_{\mathfrak{R}} (x - E[X])^2 f(x)dx = \dots = \int_{\mathfrak{R}} x^2 f(x)dx - (E[X])^2$$

La siguiente tabla resume y compara los conceptos que utilizamos, tanto cuando disponemos de datos muestrales, como cuando recurrimos al modelo teórico (en los dos casos, discreto y continuo):

Muestra	Modelo discreto	Modelo continuo
Frec. relativas= $f_i = \frac{n_i}{n}$	F. de Prob.= $P(x_i)$	F. de densidad= $f(x)$
Media muestral: $\sum_{i=1}^k f_i x_i$	Media del modelo: $\sum_{i=1}^k x_i P(x_i)$	Media del modelo: $\int_{\mathfrak{R}} x f(x) dx$
Varianza muestral: $\sum_{i=1}^k f_i (x_i - \bar{x})^2$	Varianza del modelo: $\sum_{i=1}^k (x_i - E[X])^2 P(x_i)$	Varianza del modelo: $\int_{\mathfrak{R}} (x - E[X])^2 f(x) dx$

También podemos hablar de la mediana de una variable aleatoria X , tanto en el caso discreto como en el caso continuo. La idea es sencilla: la mediana muestral era el valor que dejaba el 50% de los datos por debajo y el 50% de los datos por encima. La definición de mediana para una variable aleatoria sigue la misma filosofía: buscamos el valor que deja el 50% de la probabilidad por debajo y el 50% de la probabilidad por encima. En el caso continuo, esta idea se formaliza de manera muy sencilla (en el caso discreto, la formalización es un poco más incómoda):

Definición.- Consideramos una variable aleatoria continua X con función de densidad $f(x)$.

La **mediana** de X se define como el valor M que verifica:

$$\int_{-\infty}^M f(x) dx = \frac{1}{2}$$

4 Modelos de probabilidad más importantes

En esta sección, presentaremos los modelos de probabilidad más interesantes desde el punto de vista de las aplicaciones.

Definición.- Una **prueba de Bernoulli** es un experimento aleatorio cuyos posibles resultados son agrupados en dos conjuntos excluyentes que llamaremos éxito (E) y fracaso (F), con $P(E) = p$ y $P(F) = 1 - p$.

Esta división en éxito y fracaso puede ser algo que viene dado de manera natural o una división artificial que a nosotros nos interesa realizar. Vemos a continuación algunos ejemplos sencillos:

En el lanzamiento de una moneda podemos tomar $E = \{Cara\}$ y $F = \{Cruz\}$.

En el lanzamiento de un dado podemos tomar, por ejemplo, $E = \{1, 2\}$ y $F = \{3, 4, 5, 6\}$.

Al elegir una persona al azar en una población podemos considerar, por ejemplo, $E = \{Altura \geq 180 \text{ cm}\}$ y $F = \{Altura < 180 \text{ cm}\}$.

Al estudiar la duración de unas determinadas piezas podemos considerar, por ejemplo, $E = \{Duracion \geq 1000 \text{ horas}\}$ y $F = \{Duracion < 1000 \text{ horas}\}$.

Las pruebas de Bernoulli generan diferentes modelos de probabilidad, algunos de ellos muy interesantes y muy utilizados. Estudiamos algunos de ellos a continuación.

Definición.- Realizamos una prueba de Bernoulli con $P(E) = p$. El **modelo de Bernoulli** de parámetro p , que representaremos abreviadamente por $Bern(p)$, es el modelo de probabilidad de la variable aleatoria que obtenemos al codificar el éxito con *uno* y el fracaso con *cero*:

$$X = \begin{cases} 1 & \text{(si obtenemos éxito) con probabilidad } p \\ 0 & \text{(si obtenemos fracaso) con probabilidad } 1 - p \end{cases}$$

Sus probabilidades son:

$$P(X = 0) = 1 - p \quad ; \quad P(X = 1) = p$$

De manera más compacta y más útil para algunos cálculos, podemos escribir:

$$P(X = x) = p^x(1 - p)^{1-x} \quad \text{para } x = 0, 1.$$

El modelo de Bernoulli es el que utilizaremos cada vez que queremos estudiar la proporción de veces, p , que ocurre un determinado suceso (éxito). Ejemplos:

Proporción de veces que obtendríamos cara al lanzar una moneda muchas veces.

Proporción de personas con altura superior a 180 cm en una población.

Proporción de piezas con duración superior a 1000 horas en una gran remesa.

La esperanza y la varianza de una distribución de Bernoulli son inmediatas de obtener (simplemente, se aplican las definiciones):

$$E[X] = p \quad ; \quad V(X) = \sum x_i^2 P(x_i) - (E[X])^2 = p - p^2 = p(1 - p)$$

Definición.- Realizamos n pruebas de Bernoulli independientes, con $P(E) = p$ en cada prueba. El **modelo binomial** con parámetros n y p , que representaremos abreviadamente por $B(n; p)$, es el modelo de probabilidad de la variable aleatoria $X =$ “Número de éxitos obtenidos en las n pruebas”. Sus probabilidades son:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{para } x = 0, 1, \dots, n$$

El hecho de que una variable aleatoria X tenga distribución $B(n; p)$ lo representaremos abreviadamente por $X \sim B(n; p)$. El mismo tipo de notación se utilizará para otros modelos de probabilidad.

La esperanza y la varianza de una variable aleatoria con distribución binomial se obtienen fácilmente utilizando las propiedades de las esperanzas y las varianzas. Para esto, definimos:

$$X_i = \begin{cases} 1 & \text{si obtenemos éxito en la prueba } i\text{-ésima} \\ 0 & \text{si obtenemos fracaso en la prueba } i\text{-ésima} \end{cases} \quad (i = 1, \dots, n)$$

De esta forma, tenemos que X_1, \dots, X_n son variables aleatorias independientes con distribución de Bernoulli y, además, $X = X_1 + \dots + X_n$. Por lo tanto:

$$E[X] = E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n] = p + \dots + p = np$$

$$V(X) = V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n) = p(1 - p) + \dots + p(1 - p) = np(1 - p)$$

Calcular probabilidades correspondientes a la distribución binomial no es nada complicado mediante una calculadora. Puede utilizarse también la tabla correspondiente a esta distribución.

El modelo de Poisson es otro de los modelos más utilizados en Probabilidad y Estadística. La forma más sencilla e intuitiva de presentarlo es como límite de la distribución binomial $B(n; p)$, cuando $n \rightarrow \infty$ y $p \rightarrow 0$. Veamos, para esto, cual es el límite de las probabilidades binomiales, cuando $n \rightarrow \infty$, $p \rightarrow 0$ y $np \rightarrow \lambda$ ($0 < \lambda < \infty$):

$$\lim \binom{n}{x} p^x (1 - p)^{n-x} = \lim \frac{n(n-1) \dots (n-x+1)}{x!} p^x (1 - p)^{n-x}$$

$$\begin{aligned}
&= \lim \frac{n(n-1)\dots(n-x+1)}{x!n^x} (np)^x (1-p)^{n-x} \\
&= \lim \frac{1}{x!} \left(\frac{n}{n} \frac{n-1}{n} \dots \frac{n-x+1}{n} \right) (np)^x \frac{(1-p)^n}{(1-p)^x} \\
&= \frac{e^{-\lambda} \lambda^x}{x!}
\end{aligned}$$

Este resultado es el que motiva la siguiente definición para el modelo de probabilidad de Poisson:

Definición.- El **modelo de Poisson** de parámetro λ ($\lambda > 0$), que representaremos abreviadamente por Poisson (λ), es el modelo que tiene las siguientes probabilidades:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{para } x = 0, 1, \dots$$

Es inmediato comprobar que, efectivamente, lo anterior es una función de masa:

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

Ya que hemos presentado el modelo de Poisson como límite del modelo binomial, es fundamental comprender en qué situaciones nos encontraremos, de manera aproximada, con el modelo de Poisson: decir que $n \rightarrow \infty$ lo entenderemos, intuitivamente, como que n es grande, y decir que $p \rightarrow 0$ lo entenderemos como que p es próximo a cero. Por tanto, cuando nos encontremos con un modelo binomial con las circunstancias indicadas, lo podremos sustituir por el correspondiente modelo de Poisson, tomando $\lambda = np$ (a título orientativo, digamos que esta sustitución puede resultar aconsejable cuando $n \geq 30$, $p \leq 0.1$ y $np \leq 10$). Veamos algunos ejemplos en los que surge, de manera natural, la distribución de Poisson como modelo de la variable aleatoria considerada:

X = “Número de erratas por página en un libro”.

X = “Número de asegurados en una compañía que han declarado algún siniestro en un año”.

En resumen, digamos que la distribución de Poisson es un modelo bastante razonable cuando estamos interesados en estudiar el número de éxitos obtenidos en un número grande de pruebas independientes de Bernoulli, y la probabilidad de éxito, cada vez que se repite la prueba, es pequeña.

Una manera informal, pero sencilla, de obtener la esperanza y la varianza de una variable aleatoria X con distribución de Poisson, consiste en recordar que la distribución de Poisson es límite del modelo binomial; de este modo, tenemos:

$$E[X] = \lim np = \lambda \quad ; \quad V(X) = \lim(np)(1-p) = \lambda$$

Como se ha indicado, el procedimiento no es riguroso, pero es cómodo para recordar el valor de la esperanza y de la varianza; aplicando directamente las definiciones obtenemos los mismos resultados, pero con mucho más trabajo.

Es sencillo calcular probabilidades correspondientes al modelo de Poisson mediante una calculadora. También se puede utilizar la tabla correspondiente.

La distribución Normal es el modelo más importante y más utilizado para variables aleatorias continuas. Su importancia proviene de que aparece (por supuesto, de forma aproximada) en muchas situaciones: medidas morfológicas en especies animales o vegetales (peso, altura, etc), mediciones en experimentos físicos y químicos, etc. En general, la distribución Normal surge siempre que los resultados de un experimento sean debidos a un conjunto muy grande de causas independientes que actúan sumando sus efectos, siendo cada efecto individual de poca importancia respecto al conjunto.

Definición.- El **modelo Normal** de parámetros μ y σ ($-\infty < \mu < \infty$ y $\sigma > 0$), que representaremos abreviadamente por $N(\mu; \sigma)$, es el modelo de probabilidad caracterizado por la función de densidad:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] \quad \text{para todo } x \in \mathfrak{R}$$

La forma de la función de densidad es la de una campana con su centro en $x = \mu$.

Hay una serie de propiedades básicas que conviene saber sobre la distribución Normal:

- a) $E[X] = \mu$.
- b) $V(X) = \sigma^2$.
- c) Es una densidad simétrica con respecto a la media μ .

Una consecuencia de esto es que, por ejemplo, $P(X < \mu - 1) = P(X > \mu + 1)$.

d) Si una variable aleatoria X tiene distribución $N(\mu; \sigma)$, entonces la variable aleatoria

$$Z = \frac{X - \mu}{\sigma}$$

tiene distribución $N(0; 1)$.

Gracias a esta propiedad podremos calcular la probabilidad de un suceso correspondiente a una variable aleatoria $X \sim N(\mu; \sigma)$ a partir de la tabla de la distribución $N(0; 1)$; ya veremos con un ejemplo como se hace esto.

e) La distribución $B(n; p)$ tiende a la distribución Normal, cuando $n \rightarrow \infty$ (y p está fijo).

Por tanto, cuando estemos trabajando con un modelo binomial $B(n; p)$ con n grande, lo podremos sustituir por el correspondiente modelo Normal, tomando $\mu = np$ y $\sigma = \sqrt{np(1-p)}$ (a título orientativo, digamos que esta sustitución puede resultar aconsejable cuando $n \geq 30$ y $0'1 < p < 0'9$).

Para darse cuenta de la importancia práctica de esta sustitución, podemos considerar una variable aleatoria $X \sim B(n = 100; p = 0'3)$ e intentar calcular directamente $P(X > 40)$.

f) Si $X_1 \sim N(\mu_1; \sigma_1), \dots, X_n \sim N(\mu_n; \sigma_n)$ y son independientes, entonces:

$$\begin{aligned} X_1 + \dots + X_n &\sim N\left(\mu = \mu_1 + \dots + \mu_n; \sigma = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}\right) \\ X_1 - X_2 &\sim N\left(\mu = \mu_1 - \mu_2; \sigma = \sqrt{\sigma_1^2 + \sigma_2^2}\right) \end{aligned}$$

Ejemplo.- Vamos a calcular la probabilidad de que la variable aleatoria X tome valores entre -1 y 7, si su distribución es $N(\mu = 5; \sigma = 4)$.

Para poder obtener esta probabilidad, vamos a utilizar las propiedades anteriores para transformar el cálculo de $P(-1 \leq X \leq 7)$ en algo que podamos obtener utilizando las tablas de la distribución $N(0; 1)$. Estas tablas nos dan probabilidades del tipo $P(Z > z)$ para $z > 0$. Tenemos:

$$\begin{aligned} P(-1 \leq X \leq 7) &= P\left(\frac{-1 - 5}{4} \leq \frac{X - 5}{4} \leq \frac{7 - 5}{4}\right) \\ &= P(-1'5 \leq Z \leq 0'5) \quad (\text{por la propiedad d)}) \\ &= P(Z \geq -1'5) - P(Z \geq 0'5) \end{aligned}$$

$$\begin{aligned}
&= P(Z \leq 1'5) - P(Z \geq 0'5) \quad \text{por la propiedad c))} \\
&= 1 - P(Z \geq 1'5) - P(Z \geq 0'5) \\
&= 1 - 0'0668 - 0'3085 = 0'6247 \quad \text{(utilizando las tablas)} \bullet
\end{aligned}$$

La distribución exponencial es un modelo de probabilidad sencillo que se utiliza muy a menudo para modelizar tiempos de vida de seres vivos, tiempos de vida útil de piezas,...

Definición.- El **modelo exponencial** de parámetro λ ($\lambda > 0$), que representaremos abreviadamente por $\text{Exp}(\lambda)$, es el modelo de probabilidad caracterizado por la función de densidad:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x > 0 \\ 0 & \text{en el resto} \end{cases}$$

Utilizando las definiciones y la integración por partes, obtendríamos:

$$E[X] = \frac{1}{\lambda} \quad ; \quad V(X) = \frac{1}{\lambda^2}$$

5 Muestreo aleatorio y estadísticos

El objetivo de la **Inferencia Estadística** es obtener conclusiones sobre alguna característica cuantitativa de una población a partir de los datos obtenidos en una muestra. Para poder hacer esto de una forma objetiva y científica necesitamos modelizar, tanto la característica que queremos estudiar, como la muestra con los datos que nos suministrarán la información.

Empezaremos con la modelización de la característica.

Consideraremos la característica X que nos interesa estudiar en una población como una variable aleatoria, discreta o continua, según como sea la característica en cuestión. Veamos algunos ejemplos:

Ejemplo 1.- Consideramos una moneda que no sabemos si está equilibrada o no. Queremos obtener información sobre la probabilidad de obtener cara (es decir, sobre la proporción de caras que obtendríamos en una gran cantidad de lanzamientos). Llamaremos p a la probabilidad desconocida de cara. Si lanzamos la moneda n veces para obtener información sobre p y anotamos los resultados, dispondremos de una sucesión de caras y cruces. Si codificamos las caras con un uno y las cruces con un cero (como se hace

habitualmente con los éxitos y los fracasos), dispondremos de un conjunto (X_1, \dots, X_n) de unos y ceros procedentes de la variable aleatoria:

$$X = \left\{ \begin{array}{ll} 1 & \text{(si sale cara) con probabilidad } p \\ 0 & \text{(si sale cruz) con probabilidad } 1 - p \end{array} \right\} \sim \text{Bernoulli}(p)$$

Destaquemos que el modelo de probabilidad que utilizamos para X es conocido, pero nos falta por conocer el valor del parámetro p .

Este tipo de modelización será el que utilizemos siempre que queramos estudiar una probabilidad, una proporción o un porcentaje.

Ejemplo 2.- En una fábrica se está ensayando una nueva fibra sintética, y se quiere estudiar la resistencia a la rotura de las cuerdas fabricadas con esta nueva fibra. Para poder abordar este estudio, necesitaremos datos, para lo cual medimos la resistencia de una serie de cuerdas y anotamos los resultados.

La característica o variable aleatoria que queremos estudiar en este caso es $X = \text{“Resistencia a la rotura”}$ y puede ser modelizada mediante una $N(\mu; \sigma)$. En definitiva, dispondremos de un conjunto (X_1, \dots, X_n) de datos experimentales procedentes de la variable aleatoria:

$$X = \text{“Resistencia a la rotura”} \sim N(\mu; \sigma)$$

Destaquemos que, nuevamente, el modelo de probabilidad que utilizamos para X es conocido, pero nos falta por conocer el valor de los parámetros μ y σ .

Estos ejemplos, y otros similares, se pueden plantear de un modo general:

Definición.- Consideraremos que la característica X que se quiere estudiar en una población es una variable aleatoria (discreta o continua) que puede ser modelizada por una función de probabilidad $P_\theta(x)$ (en el caso discreto), o por una función de densidad $f_\theta(x)$ (en el caso continuo), donde θ es el nombre genérico que daremos al parámetro o parámetros que identifican el modelo de probabilidad con el que estamos trabajando.

El modelo de probabilidad es conocido, pero nos falta por conocer el valor del parámetro θ .

Pasemos ahora a la modelización de la muestra.

Para poder decir algo sensato y objetivo sobre el valor desconocido del parámetro θ , necesitamos datos procedentes de la población que se está estudiando. Estos datos constituirán una muestra, es decir, un conjunto de valores (X_1, \dots, X_n) procedentes de la característica o variable aleatoria X que se desea estudiar. Ahora bien, para poder obtener conclusiones sobre

la población a partir de los datos, esta muestra tiene que cumplir algunos requisitos. Estos requisitos son los que se definen y explican a continuación:

Definición.- Una **muestra aleatoria** (o muestra aleatoria simple) de tamaño n de una característica X de una población modelizada con una función de masa $P_\theta(x)$ (en el caso discreto) o con una función de densidad $f_\theta(x)$ (en el caso continuo) es un conjunto de observaciones (X_1, \dots, X_n) donde:

1. El modelo de probabilidad de cada observación X_i es el modelo de la característica X que estamos estudiando en la población. Intuitivamente, esto significa que cada observación X_i es **representativa** de la población de procedencia, de modo que los valores más frecuentes en la población aparecerán con más frecuencia en la muestra.
2. Las observaciones X_1, \dots, X_n son **independientes**. Intuitivamente, esto significa una de las dos siguientes cosas:
 - (a) El muestreo se realiza con reemplazamiento.
 - (b) El muestreo se realiza sin reemplazamiento, pero el tamaño muestral es pequeño en comparación con el tamaño de la población, de modo que, en la práctica, es como si fuera con reemplazamiento.

Desde un punto de vista técnico, todo lo anterior se puede resumir de la siguiente forma:

Función de probabilidad de la muestra (caso discreto):

$$P_\theta(x_1, \dots, x_n) = P_\theta(x_1) \dots P_\theta(x_n)$$

Función de densidad de la muestra (caso continuo):

$$f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \dots f_\theta(x_n)$$

Finalmente, indiquemos que hay otros tipos de muestreo, aunque aquí vamos a limitarnos al muestreo aleatorio.

Definición.- Un **estadístico** es una función T de la muestra aleatoria (X_1, \dots, X_n) , que utilizaremos como resumen de esa muestra.

Algunos de los estadísticos más utilizados en todo tipo de situaciones son los siguientes:

$$\text{Media muestral}=\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\text{Varianza muestral}=V_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right]$$

$$\text{Cuasi-varianza muestral}=S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right]$$

Muchos de estos estadísticos ya aparecieron en la Estadística Descriptiva como resúmenes de los datos de una muestra. Sólo hay una diferencia (de tipo técnico): un estadístico puede considerarse como una variable aleatoria y en consecuencia podemos hablar de su esperanza, de su varianza, etc.

Veamos algunas propiedades importantes de estos estadísticos:

Propiedades.- Sea (X_1, \dots, X_n) una muestra aleatoria de una característica X en una población, con esperanza μ y varianza σ^2 . Entonces:

1. $E[\bar{X}] = \mu$
2. $V(\bar{X}) = \frac{\sigma^2}{n}$
3. $E[S^2] = \sigma^2$
4. $E[V_X] = \frac{n-1}{n}\sigma^2$

Las técnicas empleadas en la Inferencia Estadística se agrupan en tres grandes bloques, dependiendo de la naturaleza del problema que intentemos resolver y del tipo de solución que demos: **estimación puntual, intervalos de confianza** y **contraste de hipótesis**. Estos tres grandes bloques se estudiarán en los próximos temas.