

Inferencia Estadística

- **Inferencia Estadística**
 - Estimación puntual
 - Estimación por intervalos
 - Estimación bayesiana
 - Contraste de hipótesis

Inferencia Estadística: Estimación puntual

Estimación puntual

- Introducción a la inferencia estadística
- Métodos de muestreo
- La estimación puntual
- Distribución de un estimador en el muestreo
- Propiedades de los estimadores
- Estimadores de máxima verosimilitud

Introducción a la inferencia estadística

- La creación de modelos probabilísticos es un caso típico de razonamiento deductivo donde se generan las hipótesis generales sobre el mecanismo que origina los datos, generando así las distribuciones de probabilidades que originan los datos:
 - Por ejemplo definimos e un proceso de Bernoulli la variable binomial como: $y =$ número de elementos defectuosos al observar n observaciones. Ahora suponemos que hemos realizado n observaciones de las cuales r son defectuosas y $n-r$ son aceptables (da igual el orden por la hipótesis de independencia). Con estas hipótesis dedujimos en capítulos anteriores la distribución de probabilidad binomial (razonamiento deductivo), y así con resto de distribuciones estudiadas anteriormente.
- El procedimiento inverso se realiza mediante la inferencia estadística, i.e. mediante las frecuencias observadas de una variable, extraer o inferir el modelo probabilístico que han generado los datos mostrando esas frecuencias (razonamiento inductivo).

Introducción a la inferencia estadística

- Existen muchos tipos de inferencia estadística:
 - Según el objetivo del estudio: muestreo frente a diseño.
 - **Describir** variables y sus relaciones entonces se utilizan técnicas de muestreo.
 - **Contrastar** relaciones entre variables y **predecir** valores futuros se utilizan técnicas de diseño experimental (se fijan valores de cierta variables y se miden la respuesta que inducen otras).

Introducción a la inferencia estadística

- Existen muchos tipos de inferencia estadística:
 - Por el método utilizado: métodos paramétricos v.s. no paramétricos.
 - **Paramétrico:** se supone que los datos provienen de una cierta distribución y se tienen muestras para estimar los parámetros de la misma.
 - **No paramétrico:** supones aspectos generales de la distribución (continua simétrica, etc.) y tratan de estimar o contrastar su estructura. Generalmente se estiman su forma mediante el suavizado los histogramas de los datos muestrales.

Introducción a la inferencia estadística

- Existen muchos tipos de inferencia estadística:
 - Por la información considerada: enfoque clásico v.s. bayesiano.
 - **Clásico:** los parámetros son cantidades fijas desconocidas (sin información sobre ellos), y la inferencia utiliza solo la información de los datos muestrales.
 - **Bayesiano:** considera los parámetros como variables aleatorias y permite introducir información adicional sobre los mismos a través de una probabilidad a priori.

Métodos de muestreo: muestra y población

- **Población:** conjunto homogéneo de elementos en los que se estudia una característica dada. Normalmente no es posible estudiar toda la población:
 - Destrucción de los elementos: ej. estudiar la tensión de rotura de cables.
 - Lo elementos pueden existir conceptualmente, pero no en la realidad: ej. Población de piezas defectuosas que producirá una máquina.
 - Inviabile económicamente estudiar toda la población.
 - El estudio llevaría tanto tiempo que sería impracticable.
- Se suele elegir un conjunto representativo que es la **muestra**, y si esta se selecciona bien podemos obtener una información similar de la población.
- La clave es seccionar la muestra **representativa** de la población.

Métodos de muestreo: muestreo aleatorio simple

- Una muestra es aleatoria simple si (m.a.s.):
 - Cada elemento de la población tiene la misma probabilidad de ser elegido.
 - Las observaciones se realizan con reemplazamiento (población idéntica en todas las extracciones).
- La primera condición asegura representatividad de la muestra (si A esta en el 20% y todos los elementos tienen idéntica probabilidad de ser seleccionados, la muestra tendrá un 20% también).
- La segunda se impone por simplicidad.

Métodos de muestreo: muestreo aleatorio simple

- En una muestra aleatoria cada observación tiene la distribución de probabilidad de la población.
- Sea la muestra observada $X'=(x_1, \dots, x_n)$, donde x_i representa el valor de x en el elemento i -ésimo.
- Llamamos f_1, \dots, f_n a las funciones de densidad de esas variables que verifican en el muestreo aleatorio simple que $f_1 = \dots = f_n = f$.
- Como las observaciones son independientes en una muestra aleatoria simple, entonces la distribución conjunta de la muestra se puede poner $f_c(x_1, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n) = f(x_1) f(x_2) \dots f(x_n)$.

Métodos de muestreo: otros tipos de muestreo

- Muestreo **estratificado**: El método anterior se utiliza cuando la población es homogénea. Cuando se tiene información heterogénea de la población hay que dividir la población en estratos o clases, realizando un muestreo aleatorio simple dentro de cada estrato (ej. Encuestas de opinión, que se divide por sexo, edad, profesión, etc.).
- Supongamos k estratos de tamaños N_1, \dots, N_k , con $N = N_1 + \dots + N_k$.
- La muestra que tomemos debe garantizar la presencia adecuada de cada estrato.
- Existen criterios básicos para dividir el tamaño total de muestra (n) entre los estratos (n_i):
 - Proporcional: $n_i = n(N_i/N)$.
 - Proporcional a la variabilidad del estrato: los estratos variables están más representados. Si σ_i es la variabilidad dele estrato i , entonces:

$$n_i = n (\sigma_i N_i) / (\sum_{i=1}^k \sigma_i N_i)$$

Métodos de muestreo: otros tipos de muestreo

- Muestreo por **conglomerados**: Hay situaciones en las cuales donde ni el muestreo aleatorio simple ni el estratificado pueden darse. En estos casos la población se encuentra agrupada en conglomerados, cuyo número se conoce.
- Ej. La población se distribuye en provincias, los habitantes de provincias en ciudades, etc.
- Si se supones los conglomerados independientes se pueden analizar con la metodología anterior.

Métodos de muestreo: otros tipos de muestreo

- Muestreo **sistemático**: Cuando los elementos están ordenados en listas. Supongamos que queremos una muestra n de una población N . Calculamos $k=N/n$. Se coge un elemento entre los primeros k , supongamos que el orden del elegido es n_1 , tomamos a continuación los elementos n_1+k , n_1+2k , etc., hasta completar la muestra, es decir n veces (siendo k el número de grupos de tamaño n en la población).
 - Si el orden en la lista es al azar este procedimiento es equivalente al muestreo aleatorio simple.
 - Si el orden en la lista es de la forma que elementos cercanos son más similares que los más alejados, entonces se puede demostrar que este procedimiento cubre más homogéneamente toda la población, siendo más preciso que el muestreo aleatorio simple.

La estimación puntual: fundamentos

- **Supongamos que se observa una muestra aleatoria simple de una variable aleatoria x siguiendo una distribución conocida** como las que hemos estudiado: distribución normal, Poisson, etc.
- Lo que **no conocemos son los parámetros** de esas distribuciones conocidas.
- A las cantidades que estiman los parámetros de la distribución de la población a través de datos muestrales se le llaman **estimadores estadísticos**.
- ¿Cómo estimamos esos parámetros de los datos muestrales recogidos?

La estimación puntual: fundamentos

- En primera aproximación supondremos que **no tenemos ningún tipo de información del parámetro a ajustar, ϑ** , de la distribución supuesta.
- **Si hubiese algún tipo de evidencia sobre el parámetro a estimar se utiliza el enfoque bayesiano** (más adelante).
- Así el enfoque que vamos a ver ahora es el paramétrico, que dependiendo del tipo de variable a estudiar supondrá un modelo u otro a ajustar sus parámetros.

La estimación puntual: la identificación del modelo

- La primera operación a realizar con la muestra en un análisis descriptivo, para así ver si el modelo que consideramos es consistente con la muestra.
- Si tenemos muestra pequeñas (menos que 30), es más complicado y se suelen hacer ciertos tipos de gráficos que nos sacan de dudas.
- Para chequear visualmente si una muestra pequeña la podemos asociar a una distribución de Poisson:
 - Si **siguen una distribución de Poisson** entonces $E[f_{ob}(x)] = nP(x) = n \frac{\lambda^x}{x!} e^{-\lambda}$, siendo n el tamaño de la muestra.
 - Si sacamos logaritmos neperianos: $\ln E[f_{ob}(x)] = \ln n - \lambda + x \ln \lambda - \ln x!$, por lo tanto $\ln E[f_{ob}(x)] + \ln x! = \ln n - \lambda + x \ln \lambda = A + xB$.
 - Por tanto si dibujamos $\ln f_{ob}(x) + \ln x!$ respecto de x tiene que salir casi una recta si los valores esperados se distribuyen según una distribución de Poisson (ver ejemplo numérico en el libro).
 - La recta debería tener una pendiente $\ln \lambda$ y ordenada en el origen $\ln n - \lambda$.

La estimación puntual: la identificación del modelo

- Existen otros métodos para comprobar otras distribuciones, como por ejemplo para la distribución normal:
 - Que se puede utilizar un papel probabilístico normal (ver ejemplo en el libro) para dibujar los datos. Si no se ajustan a una recta, los datos no se distribuyen según una normal.
 - También se pueden usar los gráficos Q-Q plots con la misma idea anterior, que ya hemos visto anteriormente.
- En general los gráficos Q-Q plots se pueden utilizar con cualquier distribución de probabilidad, como ya comentamos anteriormente.

La estimación puntual: el método de los momentos

- Es el primer método que se utilizó para obtener el estimador de un parámetro de una distribución dada (formalizado por K. Pearson).
- Se toma como estimador de la varianza de la población la varianza de la muestra, de la media de la población la media muestral, y así sucesivamente con todos los momentos que se quieran incluir en la estimación paramétrica.
- Se trata de estimar un vector de parámetros $\underline{\vartheta} = (\vartheta_1, \dots, \vartheta_k)$, cuyos componentes se pueden expresar en función de los k momentos de la población, m_k , siendo $\vartheta_1 = g_1(m_1, \dots, m_k)$, \dots , $\vartheta_k = g_k(m_1, \dots, m_k)$.

La estimación puntual: el método de los momentos

- Así estimamos los correspondientes momentos muestrales, $\hat{m}_1, \dots, \hat{m}_k$, sustituyéndolos en el sistema de ecuaciones anteriores, obteniendo los parámetros estimados en la población: $\hat{\vartheta}_1, \dots, \hat{\vartheta}_k$.
- Es muy importante estimar cual es la bondad de ajuste de estos estimadores de los parámetros y sus propiedades deseables. Esto es lo que vamos a ver a continuación.

Distribución de un estimador en el muestreo: concepto

- **Podemos ver el estimador como una variable aleatoria**, cuyo valor cambia de muestra en muestra.
- Por ejemplo supongamos la muestra (2, 4, 9, 1) de una distribución uniforme en el intervalo (0, b), para estimar b de la muestra (el valor esperado de una distribución uniforme en a,b, es $a+b/2$):
 - $E[x]=(0+b)/2$, por tanto $\hat{b} = 2\bar{x}=2(2+4+9+1/4)=8$, este estimador no es el más preciso, ya que pudiéramos elegir el máximo, 9. Además si tenemos otra muestra cambia.
- Consideremos una población de la que se toman muestras con remplazamiento de tamaño n y calculamos en cada muestra la media \bar{x} .

Distribución de un estimador en el muestreo: concepto

- Así si tomamos k muestras obtendremos en general k valores de las medias muestrales: $\bar{x}_1, \dots, \bar{x}_k$. Si k es muy grande tendiendo a infinito los valores los valores \bar{x}_i tendrán una distribución que llamaremos **distribución muestral de la media** en el muestreo.
- Esta distribución en el muestreo de un estadístico depende de:
 - La población base.
 - El tamaño de la muestra n .

Distribución de un estimador en el muestreo: concepto

- Recordemos que el estudio de determinadas características de una población se efectúa a través de diversas muestras que pueden extraerse de ella (no olvidar).
- En general el muestreo se puede realizar con o sin reposición.
- La población de partida puede ser infinita o finita.
- Una población finita en la que se efectúa muestreo con reposición podría considerarse infinita, aunque también una población muy grande puede considerarse como infinita.
- Aquí vamos a limitarnos a una población de partida infinita o a muestreo con reposición.

Distribución de un estimador en el muestreo: concepto

- Si consideramos todas las posibles muestras de tamaño n en una población, para cada muestra se puede calcular estadísticos como la media, desviación típica, proporción, etc., que variarán de una muestra a otra.
- Así obtenemos una distribución del estadístico que consideremos que se es lo que se llama **distribución muestral**, en general.
- Un **estadístico** (refiriéndose a datos muestrales, i.e. estadístico muestral) es una medida cuantitativa, derivada de un conjunto de datos de una muestra, con el objetivo de estimar o inferir características de una población o modelo estadístico.

Distribución de un estimador en el muestreo: concepto

- En general podemos definir un **Estadístico** como una función de los valores de la muestra. Es una variable aleatoria, cuyos valores dependen de la muestra seleccionada.
- Su distribución de probabilidad, se conoce como **Distribución muestral del estadístico**.
- Por ejemplo para el estadístico media de las diferentes muestras de una población se obtiene la distribución muestral de la media, para la varianza la distribución muestral de la varianza.

Distribución de un estimador en el muestreo: distribución en el muestreo de una proporción

- Supongamos una población donde observamos la presencia o no de un atributo. Y sea p la proporción desconocida de elementos con dicho atributo en la población.
- La distribución del muestreo del estimador de \hat{p} en la muestra, viene determinada por la distribución binomial:
 - $P\left(\hat{p} = \frac{r}{n}\right) = PB(r) = \binom{n}{r} p^r (1-p)^{n-r}, r=0, 1, \dots, n$
- Por lo tanto la **probabilidad de que la proporción** en la muestra sea r/n es igual a la probabilidad de obtener r elementos con una característica determinada en una muestra de tamaño n , que es directamente la **distribución binomial**.

Distribución de un estimador en el muestreo: distribución en el muestreo de una proporción

- Así las propiedades de la distribución en el muestreo del estimador \hat{p} vendrán dadas por la propiedades del valor esperado y la varianza del estimador de la proporción:
 - $\mathbf{E}[\hat{p}] = E[r/n] = (1/n)E[r] = np/n = \mathbf{p}$,
 - (recordar que en una binomial $E[r] = np$).
 - $\mathbf{Var}[\hat{p}] = E[r/n] = (1/n)^2 \text{Var}[r] = \mathbf{pq/n}$,
 - (recordar que en una binomial $\text{Var}[r] = npq$).

Distribución de un estimador en el muestreo: distribución en el muestreo de una proporción

- Cuando k es grande, la distribución de muestreo de \hat{p} será aproximadamente normal con la media y varianza de las dos expresiones anteriores, ya que es un caso particular de la distribución muestral de una media, ya que \hat{p} se calcula por: $\hat{p} = (x_1 + \dots + x_n) / n$ y entonces se puede aplicar aplica el TCL.
- Cada x_i toma el valor 1 si el elemento tiene el atributo estudiado, y 0 en cualquier otro caso.
- **Así \hat{p} es la media muestral de las variables de Bernoulli, x_i .**

Distribución de un estimador en el muestreo: distribución muestral de la media

- Para calcular la distribución muestral de la media tenemos que tener en cuenta que cada muestra de tamaño n que podemos extraer de una población proporciona una media.
 - Podemos considerar cada una de estas medias como valores de una variable aleatoria y podemos estudiar su distribución que llamaremos **distribución muestral de las medias**.
- Vamos a calcular la media y varianza de la distribución muestral de la media, en el caso general en el que la variable aleatoria x tiene media μ y varianza σ^2 .

Distribución de un estimador en el muestreo: distribución muestral de la media

- Para el cálculo de distribución muestral de la media suponemos que cada muestra es de tamaño n , y suponemos que todas las variables x_i de una muestra aleatoria simple tiene la misma distribución de la población.
- Así el valor esperado la distribución muestral de las medias y su varianza vienen determinados por las expresiones:
 - $E[\bar{x}] = E[(1/n)\sum x_i] = 1/n \sum E[x_i] = 1/n \sum \mu = \mu.$
 - $\text{Var}[\bar{x}] = (1/n)^2 \sum \text{Var}[x_i] = n\sigma^2 / n^2 = \sigma^2 / n.$
- Hemos aplicado el hecho de que para la variable aleatoria x_i se cumple que $E[x_i] = \mu$ y $\text{Var}[x_i] = \sigma^2$, como hemos dicho antes.

Distribución de un estimador en el muestreo: distribución muestral de la media

- Resumiendo, al tomar una muestra de tamaño n de una variable con media μ y varianza σ^2 y distribución cualquiera, la distribución muestral de la media verifica que $E[\bar{x}] = \mu$ y $\text{Var}[\bar{x}] = \sigma^2 / n$.
- En el caso de la distribución de muestreo de la proporción $E[\hat{p}] = p$ y $\text{Var}[\hat{p}] = pq/n$, es un caso especial de este que acabamos de ver con media p y varianza pq .
- Si tenemos una población normal $N(\mu, \sigma)$ y extraemos de ella muestras de tamaño n , la distribución muestral de medias sigue también una distribución normal $N(\mu, \sigma/(n)^{0.5})$ ([volver](#)).
- Si la población no sigue una distribución normal pero $n > 30$, aplicando el llamado el TCL la distribución muestral de medias se aproxima también a la normal anterior.

Distribución de un estimador en el muestreo: distribución muestral de la varianza

- Para la **distribución muestral de varianzas**, podemos seguir los mismos razonamientos anteriores y suponemos de nuevo una variable aleatoria x que se observa en la muestra que tiene media μ y varianza σ^2 .
- Se puede calcular que esperanza de la distribución de varianza de la muestra es $E[s^2]=\sigma^2(n-1)/n$. En consecuencia el valor medio de s^2 es menor que σ^2 , aunque la diferencia tiende a cero a aumentar el tamaño de la muestra n .
- Se puede definir la varianza muestral corregida como $\hat{s}^2=(n/n-1) s^2$, de tal forma que $E[\hat{s}^2]=\sigma^2$.
- Estas propiedades se verifican siempre, cualquiera que sea la distribución de la variable x .
- Así se pueden calcular la distribución de cualquier estimador en el muestreo, en libro vienen más ejemplos (capítulo 7).

Propiedades de los estimadores: centrado o insesgado

- Diremos que un **estimador** $\hat{\vartheta}$ es **centrado** o insesgado para ϑ si para cualquier tamaño muestral tenemos que $E[\hat{\vartheta}] = \vartheta$.
- Cuando no es centrado se define el **sesgo** del estimador como $\text{sesgo}(\hat{\vartheta}) = E[\hat{\vartheta}] - \vartheta$.
- **Pueden existir muchos estimadores centrados para un parámetro:**
 - Para estimar μ en una distribución cualquiera todos los estimadores del tipo siguiente son centrados: $\hat{\mu} = a_1x_1 + \dots + a_nx_n$ con $\sum a_i = 1$.
- Anteriormente hemos comprobado que \bar{x} (y como caso particular el estimador de la proporción \hat{p}) es siempre centrado para estimar μ .

Propiedades de los estimadores: centrado o insesgado

- También hemos visto que s^2 no es centrado para estimar σ^2 (recordar que lo corregimos).
- Una ventaja fundamental de los **estimadores centrados** es que **los podemos combinar para obtener nuevos estimadores centrados**:
 - Si tenemos dos muestras independientes y calculamos en cada una de ellas un estimador centrado $\hat{\vartheta}_i$ para un parámetro determinado, cualquier estimador del tipo $\hat{\vartheta}_t = a_1\hat{\vartheta}_1 + a_2\hat{\vartheta}_2$, con $a_1 + a_2 = 1$, es un estimador centrado.
- Los **estimadores centrados no tienen que ser los mejores**, alguna veces es preferible tener uno sesgado con poca varianza, que uno no sesgado pero con mucha varianza (**eficiencia de estimadores**).

Propiedades de los estimadores: eficiencia o precisión

- La **eficiencia** o precisión se define en función del **inverso de la varianza**: precisión $(\hat{\vartheta}) = 1/Var[\hat{\vartheta}]$.
- Diremos que un estimador $\hat{\vartheta}_2$ es más eficiente que un estimador $\hat{\vartheta}_1$ si para cualquier tamaño muestral se cumple que $Var(\hat{\vartheta}_2) \leq Var(\hat{\vartheta}_1) \Leftrightarrow Efic(\hat{\vartheta}_2) \geq Efic(\hat{\vartheta}_1)$.
- Se define la eficiencia relativa de $\hat{\vartheta}_2$ respecto a $\hat{\vartheta}_1$ al cociente entre sus eficiencias: $ER(\hat{\vartheta}_2/\hat{\vartheta}_1) = Efic(\hat{\vartheta}_2)/Efic(\hat{\vartheta}_1) = Var(\hat{\vartheta}_1)/Var(\hat{\vartheta}_2)$.
- La eficiencia de estimadores esta completamente ligada a la varianza de los mismos.
- Mirar en el libro como combinar linealmente estimadores centrados para minimizar la varianza y el ejemplo 7.4.

Propiedades de los estimadores: error cuadrático medio

- Muchas veces se nos presenta el problema de elegir entre dos estimadores uno centrado y con varianza no muy grande y otro sesgado y con varianza un poco más pequeña. En estos casos se elige aquel que tiene el menor error cuadrático medio con el parámetro que esta intentando estimar.
- Así tenemos $ECM(\vartheta) = E[(\hat{\vartheta} - \vartheta)^2]$, tomando el promedio respecto a la distribución en el muestreo del estimador.
- Se puede demostrar que $ECM(\vartheta) = [sesgo(\hat{\vartheta})]^2 + Var(\hat{\vartheta})$.

Propiedades de los estimadores: consistencia y robustez

- Cuando lo único que podemos tener son estimadores segados y no con mucha eficiencia, se le pide al estimador que sea **consistente**.
- Un estimador se dice que es consistente si cuando crece el tamaño de la muestra el estimador tiende al parámetro que se está estimando: $\lim_{n \rightarrow \infty} E[\hat{\vartheta}_n] \rightarrow \vartheta$.
- Es decir la esperanza del estimador es asintóticamente el valor del parámetro.

Propiedades de los estimadores: consistencia y robustez

- En este caso la varianza del estimador va a cero también con el tamaño de la muestra: $\lim_{n \rightarrow \infty} Var[\hat{\vartheta}_n] \rightarrow 0$.
- Un buen estimador es **robusto** para un parámetro ϑ en el modelo $f(x)$, si variando débilmente el modelo este estimador experimenta una pequeña modificación (ver ejemplo en el libro, capítulo 7 de contaminación de un modelo normal como decrece la eficiencia).

Estimadores de máxima verosimilitud: distribución conjunta de la muestra

- Los conceptos de funciones de verosimilitud se deben a Fisher, y es fundamental en inferencia estadística.
- Este concepto se define a partir de la distribución conjunta de la muestra.
- Supongamos una variable discreta x con distribución $P(x; \vartheta)$ que es conocida.
- Supongamos que tomamos muestras independientes de tamaño n , representando está por el vector \mathbf{X} .
- Así podemos definir la distribución conjunta de la muestra en función de esta variable, y para el caso de una muestra aleatoria simple tenemos:
 - $P(\mathbf{X} = \mathbf{X}_0) = P(x_1 = x_{10}, x_2 = x_{20}, \dots, x_n = x_{n0}) = P(x_{10}) \dots P(x_{n0})$
- Así conociendo la distribución $P(x; \vartheta)$ podemos calcular fácilmente la probabilidad de cualquier muestra.

Estimadores de máxima verosimilitud: distribución conjunta de la muestra

- En el **caso continuo** (función de densidad $f(x; \vartheta)$), la probabilidad del intervalo $x_1 - 1/2, x_1 + 1/2$, la podemos aproximar por el rectángulo de altura $f(x_i)$ y base unidad:
 - $P(x_i) = f(x_i) \cdot 1$
- Por tanto la probabilidad de la muestra aleatoria simple:
 - $P(x_1, \dots, x_n) = \prod f(x_i)$
- Así la función de densidad conjunta de la muestra $f(x_1, \dots, x_n)$ se interpreta como la probabilidad de obtener los valores muestrales $x_1 \pm 0,5, \dots, x_n \pm 0,5$.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Sea una variable aleatoria continua x con función de densidad que $f(x | \vartheta)$ para indicar que depende de un vector de parámetros ϑ . Es decir dado que conozco ϑ , representa cual es la función de densidad de la variable aleatoria x .
- Si tenemos una muestra aleatoria simple $\mathbf{X} = (x_1, \dots, x_n)$, entonces la función de densidad conjunta de la muestra es:
 - $f(\mathbf{X} | \vartheta) = \prod f(x_i | \vartheta)$
- Es decir cuando conozco ϑ la expresión anterior determina la probabilidad de aparición de cada muestra.

Estimadores de máxima verosimilitud: la función de verosimilitud

- En inferencia para un problema de estimación se conoce un valor particular de una muestra \mathbf{X} , siendo desconocido el parámetro ϑ .
- Así si sustituimos \mathbf{X} por el valor observado de una muestra, $\mathbf{X}_0 = (x_{10}, \dots, x_{n0})$, entonces la función $f(\mathbf{X}_0 | \vartheta)$ puede ser vista como una función del parámetro.
- Es decir $f(\mathbf{X}_0 | \vartheta)$ puede ser visto y proporciona, para cada valor de ϑ , la probabilidad de obtener el valor muestral \mathbf{X}_0 para ese ϑ .

Estimadores de máxima verosimilitud: la función de verosimilitud

- Así esta nueva función que obtenemos cuando variamos ϑ , mientras mantenemos \mathbf{X}_0 fijo (no variamos la muestra), define la **función de verosimilitud**, $\ell(\vartheta | \mathbf{X})$ (dado que conozco la muestra como varia la probabilidad en función del parámetro ϑ)
 - $\ell(\vartheta | \mathbf{X})$, o $\ell(\vartheta)$: Es decir $\ell(\vartheta | \mathbf{X}) = \ell(\vartheta) = f(\mathbf{X}_0 | \vartheta)$, con \mathbf{X}_0 fijo y ϑ variable.
- La óptica cambia, en vez de tener un parámetro fijo ϑ y calcular para ese parámetro la probabilidad de obtener distintas muestras \mathbf{X} , lo que fijamos es una determinada muestra \mathbf{X}_0 y estimamos que valor del parámetro ϑ hace más verosímil la muestra que se observa \mathbf{X}_0 .

Estimadores de máxima verosimilitud: la función de verosimilitud

- Este enfoque cambia completamente la forma de la función.
- Si tenemos una variable x que distribuye según una Poisson:
 - $P(x = r) = \frac{\lambda^r}{r!} e^{-\lambda}$, $r = 0, 1, 2, \dots$, y observamos el valor de la muestra $x=5$, entonces $\ell(\lambda) = \frac{\lambda^5}{5!} e^{-\lambda}$, es la función de verosimilitud para una muestra de un solo valor de $x=5$.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Esta función $\ell(\lambda)$ es continua en λ y proporcional a la probabilidad de observar $x=5$ para cada valor posible de λ .
- El valor de la verosimilitud no es único: $\ell(\vartheta_1) = f(\mathbf{X}_0|\vartheta_1) > f(\mathbf{X}_0|\vartheta_2) = \ell(\vartheta_2)$.
- Esto quiere decir que a la vista de los datos muestrales el valor del parámetro ϑ_1 es más verosímil que el valor del parámetro ϑ_2 ya que la probabilidad de obtener la muestra observada \mathbf{X}_0 es mayor con ϑ_1 que con ϑ_2 .

Estimadores de máxima verosimilitud: la función de verosimilitud

- Observar que la verosimilitud tiene unidades, las de la variable x , entonces la diferencia de verosimilitudes no tiene sentido ya que varía arbitrariamente en función de las unidades de la variable x .
- Por lo tanto para comparar verosimilitudes lo mejor es el cociente de las mismas ya que este es invariante frente a las diferentes unidades de la variable:
 - $\ell(\vartheta_1 | \mathbf{X})/\ell(\vartheta_2 | \mathbf{X})$, este cociente es invariante hacia cambios de escalas en la variable x que se está observando.
 - El cociente $\ell(\vartheta_1)/\ell(\vartheta_2)$ se puede sustituir por la diferencia de logaritmos:
 $\ln\ell(\vartheta_1) - \ln\ell(\vartheta_2)$
- Así se puede definir la **función soporte** por el logaritmo de la verosimilitud: $L(\vartheta) = \ln\ell(\vartheta)$.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Se define como la discriminación contenida en la muestra \mathbf{X} entre ϑ_1 y ϑ_2 a la siguiente expresión (diferencia de soporte de ambos valores):
$$L(\vartheta_2) - L(\vartheta_1) = \ln \ell(\vartheta_2) - \ln \ell(\vartheta_1).$$
- Si ϑ es un parámetro cuyos valores posibles pertenecen a un intervalo ϑ_1 y ϑ_2 , llamaremos **discriminación relativa** entre ϑ_2 y ϑ_1 a:
 - $L(\vartheta_2) - L(\vartheta_1) / \vartheta_2 - \vartheta_1 = \ln \ell(\vartheta_2) - \ln \ell(\vartheta_1) / \vartheta_2 - \vartheta_1.$
- En el límite cuando $\vartheta_2 \rightarrow \vartheta_1$ obtenemos la tasa de discriminación para la muestra \mathbf{X} respecto el parámetro ϑ valorada en el punto ϑ_1
- $d(\vartheta_1) = \lim_{\vartheta_2 \rightarrow \vartheta_1} \frac{L(\vartheta_2) - L(\vartheta_1)}{\vartheta_2 - \vartheta_1} = \left. \frac{dL(\vartheta)}{d\vartheta} \right|_{\vartheta = \vartheta_1}$, fue introducida por Fisher, que la denominó “Score”.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Si este “score” cumple que $d(\vartheta_1) > 0$, la verosimilitud aumenta para valores superiores a ϑ_1 ,
 - es decir, la muestra tiene mayor probabilidad de ocurrir con valores mayores que ϑ_1 ,
- Mientras que si $d(\vartheta_1) < 0$ el razonamiento es el contrario, la verosimilitud aumenta para valores inferiores de ϑ_1 ,
 - es decir, la muestra tiene mayor probabilidad de ocurrir con valores menores que ϑ_1 .

Estimadores de máxima verosimilitud: la función de verosimilitud

➤ Resumiendo:

1. La función de verosimilitud es la herramienta básica que nos permite juzgar la compatibilidad entre los valores muestrales observados y los posibles valores del parámetro de la distribución de probabilidad.
2. Si queremos comparar dos posibles valores del parámetro, ϑ , se debe utilizar el cociente de sus verosimilitudes, y no su diferencia, ya que la diferencia depende de la escala de medida de las variables, como hemos dicho antes.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Por ejemplo para estimar el parámetro, λ , de Poisson de la muestra observada x_1, \dots, x_n hacemos:

- $P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, distribución de Poisson.

- $\ell(\lambda) = \prod P(x_i | \lambda) = P(x_1 | \lambda)P(x_2 | \lambda) \dots P(x_n | \lambda) = \frac{\lambda^{\sum x_i}}{\prod x_i!} e^{-n\lambda}$

- Como el término $1/(\prod x_i!)$ es una constante la podemos eliminar y escribir la función de verosimilitud para la distribución de Poisson como:

- $\ell(\lambda) = e^{-n\lambda} \lambda^{\sum x_i} = e^{-n\lambda} \lambda^{n\bar{x}}$

- Así la función soporte será para la distribución de Poisson vendrá dada por:
 $L(\lambda) = -n\lambda + n\bar{x} \ln \lambda$ (mirar los ejercicios para el resto de distribuciones).

Estimadores de máxima verosimilitud: el método de máxima verosimilitud

- Una vez que tenemos calculada una función de verosimilitud para un vector de parámetros $\boldsymbol{\vartheta}$, $\ell(\boldsymbol{\vartheta})$, un procedimiento intuitivo para estimar los parámetros de la distribución a partir de los valores observados muestralmente es maximizar el valor del parámetro que sea más verosímil, es decir el que maximice la verosimilitud.
- Así podemos resolver el sistema de ecuaciones:
 - $\partial\ell(\boldsymbol{\vartheta})/\partial\vartheta_1 = 0, \dots, \partial\ell(\boldsymbol{\vartheta})/\partial\vartheta_p = 0$
- El valor que resuelve el sistema de ecuaciones, $\hat{\boldsymbol{\vartheta}}$, corresponderá a un máximo si el valor de la matriz hessiana de segundas derivadas en ese punto es definida negativa: $H(\hat{\boldsymbol{\vartheta}}) = (\partial^2\ell(\boldsymbol{\vartheta})/\partial\vartheta_i\partial\vartheta_j)_{\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}}$.

Estimadores de máxima verosimilitud: el método de máxima verosimilitud

- A la hora de hacer los cálculos de estimadores máximo-verosímiles (MV) se obtienen derivando la función soporte: $L(\vartheta) = \ln \ell(\vartheta)$, ya que la transformación logarítmica es monótona y por tanto tiene el mismo máximo.
- Recordemos que la derivada de la función soporte la habíamos definido como la tasa de discriminación: $d(\vartheta_1) = \lim_{\vartheta_2 \rightarrow \vartheta_1} \frac{L(\vartheta_2) - L(\vartheta_1)}{\vartheta_2 - \vartheta_1} = \left. \frac{dL(\vartheta)}{d\vartheta} \right|_{\vartheta = \vartheta_1}$, así podemos definir el estimador máximo-verosímil como aquel valor de los parámetros para los que se anulan la tasa de discriminación de la muestra.

Estimadores de máxima verosimilitud: Ejemplo

- Para estimar los parámetros de una normal de la muestra observada x_1, \dots, x_n :
 - $f(x) = 1/\sigma(2\pi)^{0.5} \exp\{-(1/2\sigma^2)(x-\mu)^2\}$, densidad de probabilidad normal.
 - $\ell(\mu, \sigma^2) = \prod f(x_i | \mu, \sigma^2) = f(x_1 | \mu, \sigma^2) f(x_2 | \mu, \sigma^2) \dots f(x_n | \mu, \sigma^2) =$
 $\prod_i (1/\sigma(2\pi)^{0.5}) \exp\{-(1/2\sigma^2)(x_i - \mu)^2\} =$
 $(1/(\sigma(2\pi)^{0.5})^n) \exp\{-(1/2\sigma^2)\sum(x_i - \mu)^2\}$.
- Así la función soporte será para la distribución normal vendrá dada por: $L(\mu, \sigma^2) = -n \ln(\sqrt{2\pi}\sigma) - (1/2\sigma^2)\sum(x_i - \mu)^2 =$
 $-(n/2) \ln(2\pi\sigma^2) - (1/2\sigma^2)\sum(x_i - \mu)^2$.
- Así ahora derivamos la función soporte de la muestra e igualamos a cero para sacar cuales son los estimadores de máxima verosimilitud para μ, σ^2 de la población.

Estimadores de máxima verosimilitud: Ejemplo

- Así tenemos $L(\mu, \sigma^2) = -(n/2) \ln(2\pi\sigma^2) - (1/2\sigma^2)\sum(x_i - \mu)^2$ y la optimizamos en función de μ y σ^2 :
- $\partial L(\mu, \sigma^2)/\partial \mu = 0 = (2/2\sigma^2)\sum(x_i - \mu) = \sum(x_i - \mu)/\sigma^2 = (n\bar{x} - n\mu)/\sigma^2$, así $\hat{\mu} = \bar{x}$, por lo tanto parece lógico que el estimador máxima verosimilitud para la media de la normal es la media aritmética de la muestra observada.
- $\partial L(\mu, \sigma^2)/\partial \sigma^2 = 0 = -(n/2) (2\pi/2\pi\sigma^2) + (2 \sum(x_i - \mu)^2)/(2\sigma^2)^2 = -(n/2) (1/\sigma^2) + \sum(x_i - \mu)^2/2\sigma^4 = -(n\sigma^2/2\sigma^4) + \sum(x_i - \mu)^2/2\sigma^4 = (-n\sigma^2 + \sum(x_i - \mu)^2)/2\sigma^4$ así despejando $\hat{\sigma}^2 = \sum(x_i - \mu)^2/n = s^2$, por lo tanto parece lógico que el estimador máxima verosimilitud para la varianza de la normal es la desviación típica de la muestra observada.

Estimadores de máxima verosimilitud: propiedades de los estimadores máximo-verosímiles

- Si $\hat{\vartheta}_{MV}$ son los estimadores de máxima verosimilitud de un modelo de una población a través de sus muestras estos cumplen las siguientes propiedades:
 - Asintóticamente centrados.
 - Asintóticamente normales.
 - Asintóticamente eficientes.
 - Suficiencia.
 - Invariancia.
 - Robustez.

Estimadores de máxima verosimilitud: propiedades de los estimadores máximo-verosímiles - invariancia

➤ Invariancia:

- Si $\hat{\vartheta}_{MV}$ es el estimador máximo verosímil de ϑ , entonces $h(\hat{\vartheta}_{MV})$ es el estimador máximo verosímil de $h(\vartheta)$.
- Ejemplo:
 - Sea x_1, \dots, x_n una muestra aleatoria simple de $x \sim N(\mu, \sigma)$.
 - Sabemos que $\hat{\mu}_{MV} = \bar{x}$, lo hemos demostrado anteriormente.
 - ¿Quiénes serán los estimadores de máxima verosimilitud para 3μ , μ^2 y $1/\mu^2$?
 - Por el principio de invariancia tenemos que:
 - $3\hat{\mu}_{MV} = 3\bar{x}$
 - $\hat{\mu}_{MV}^2 = \bar{x}^2$
 - $1/\hat{\mu}_{MV} = 1/\bar{x}$

Estimadores de máxima verosimilitud: propiedades de los estimadores máximo-verosímiles – consistencia y centrado

➤ Consistencia:

➤ Bajo ciertas condiciones generales, $\hat{\vartheta}_{MV}$ es un estimador consistente de ϑ .

➤ $\lim_{n \rightarrow \infty} E[\hat{\vartheta}_n] \rightarrow \vartheta.$

➤ $\lim_{n \rightarrow \infty} Var[\hat{\vartheta}_n] \rightarrow 0.$

➤ Asintóticamente centrado:

➤ Se verifica que el $\lim_{n \rightarrow \infty} E[\hat{\vartheta}_{MV}] = \vartheta.$

Estimadores de máxima verosimilitud: propiedades de los estimadores máximo-verosímiles – normalidad asintótica

➤ Normalidad asintótica :

➤ Para tamaños muestrales grandes, desarrollando en serie la función soporte en un entorno del estimador $\hat{\boldsymbol{\vartheta}}_{MV}$ para 2º orden:

➤
$$L(\boldsymbol{\vartheta}) \cong L(\hat{\boldsymbol{\vartheta}}_{MV}) + (1/2) \left(\frac{d^2 L[\hat{\boldsymbol{\vartheta}}_{MV}]}{d\boldsymbol{\vartheta}^2} \right) (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}_{MV})^2.$$

➤ Si llamamos $\hat{\boldsymbol{\sigma}}_{MV}^2 = \left(\frac{d^2 L[\hat{\boldsymbol{\vartheta}}_{MV}]}{d\boldsymbol{\vartheta}^2} \right)^{-1}$, entonces la verosimilitud se puede escribir: $\ell(\boldsymbol{\vartheta}) = \ell(\boldsymbol{\vartheta}|\mathbf{X}) = k \exp\left(\frac{1}{2\hat{\boldsymbol{\sigma}}_{MV}^2} (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}_{MV})^2\right)$, como el soporte es el log de la verosimilitud, la verosimilitud es la exp, la constante k es $\exp(L(\hat{\boldsymbol{\vartheta}}_{MV}))$.

➤ Así la verosimilitud tiene la forma de una normal, con media $\hat{\boldsymbol{\vartheta}}_{MV}$, y varianza $\hat{\boldsymbol{\sigma}}_{MV}^2$.

Inferencia Estadística: Estimación por intervalos

Estimación por intervalos

- Introducción
- Un ejemplo simple
- El método del pivote
- Principales estadísticos pivote
- Estimación autosuficiente de intervalos de confianza (bootstrap)

Estimación por intervalos: introducción

- Ya sabemos cómo obtener estimadores para un parámetro y cómo calcular una medida de la precisión del estimador: es decir su desviación típica en el muestreo.
- Siempre es conveniente dar junto al estimador un intervalo de valores entre los cuales deberá estar el valor del parámetro que se estima con alta probabilidad.
- Éste es el objetivo de la **estimación por intervalos** de confianza.
- Para ilustrar el problema vamos a ver como ejemplo la estimación de la media con una muestra de tamaño 25 en una población normal de desviación típica conocida e igual a 10.
- Antes de observar la muestra y calcular el estimador, \bar{x} , se pueden predecir las discrepancias esperadas entre el estimador (\bar{x}) y el parámetro (μ): el 95% de las veces: $|\bar{x} - \mu| \leq 1,96 \frac{\sigma}{\sqrt{n}} = 1,96 \frac{10}{\sqrt{25}} = 3,92$

Estimación por intervalos: introducción

- Esto lo veremos en la metodología a continuación como se calcula con detalle.
- Pero lo importante es que podemos predecir que el 95% de las veces $|\bar{x} - \mu|$ no será mayor de 3,92 unidades.
- Por tanto si observamos $\bar{x} = 40$, podemos asegurar que μ estará previsiblemente en el intervalo $40 \pm 3,92$, el 95% de las veces que saquemos una muestra del mismo tamaño.
- Ésta es la idea central de construcción de intervalos de confianza.
- Así llamaremos intervalo de confianza para el parámetro ϑ a un nivel de confianza $(1 - \alpha)$, a una expresión del tipo $\vartheta_1 \leq \vartheta \leq \vartheta_2$ donde los límites ϑ_1 y ϑ_2 ($40 - 3,92$ y $40 + 3,92$ en el ejemplo) dependen de la muestra, con $\alpha \leq 1$.

Estimación por intervalos: introducción

- Estos intervalos de confianza se calculan de manera tal que si tomamos muchas muestras, todas del mismo tamaño, y construimos un intervalo con cada muestra, podemos afirmar que el $100(1 - \alpha)\%$ (en el ejemplo 95%) de los intervalos así contruidos contendrán el verdadero valor del parámetro. En este caso en el ejemplo $\alpha = 0.05$.
- El otro $100 \alpha \%$ (en el ejemplo 5%) de los intervalos así contruidos no contendrán el valor verdadero del parámetro.

Estimación por intervalos: introducción

- La idea general del procedimiento es que si estimamos ϑ con el estimador máximo verosímil $\hat{\vartheta}_{MV}$, el error relativo de la estimación se define de la siguiente forma:
 - $\omega = \vartheta - \hat{\vartheta}_{MV} / \sigma(\hat{\vartheta}_{MV})$, donde $\sigma(\hat{\vartheta}_{MV})$ es la desviación típica asintótica de la distribución muestral del estadístico máximo-verosímil que sigue, asintóticamente una distribución normal estándar (por el TCL).

Estimación por intervalos: introducción

- Este resultado indica que, sea cual sea ϑ , podemos conocer aproximadamente la distribución del error relativo que cometeremos al estimar este parámetro por $\hat{\vartheta}_{MV}$.
- Esto no lleva al método del pivote para la estimación de intervalos de confianza que veremos en detalle después.
- Pero antes vamos a ver un ejemplo sencillo para entender el problema.

Estimación por intervalos: un ejemplo simple (para entender el problema)

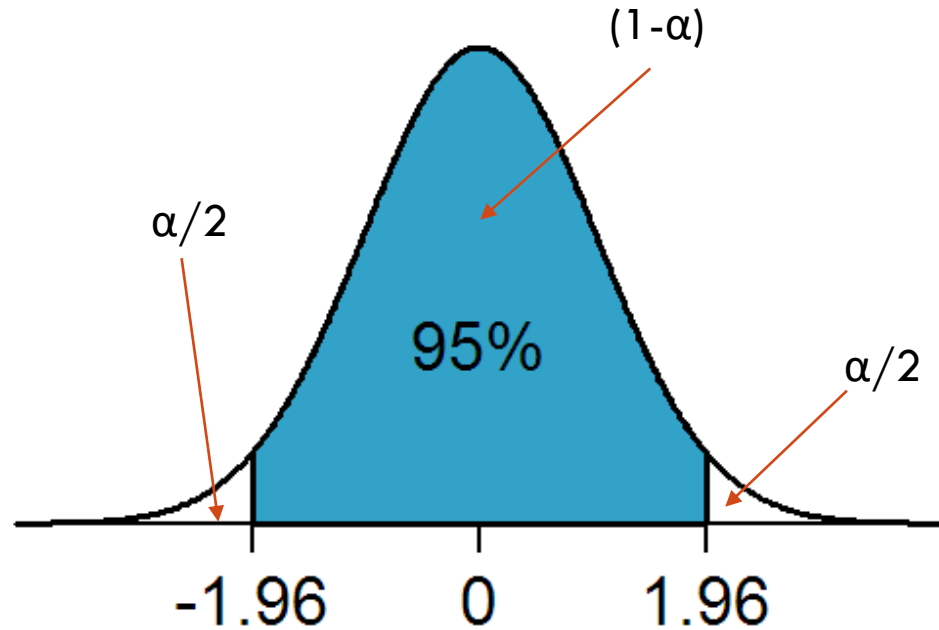
- Vamos a tomar el ejemplo que acabamos de ver con más detalle:
 - Supongamos que sabemos que tenemos una variable aleatoria que sigue una distribución $N(\mu, \sigma)$ donde el valor de varianza es conocido $\sigma^2 = \sigma_0^2$.
 - El objetivo es, dada una m.a.s. de tamaño n de la variable obtener un intervalo a un nivel de confianza del 95 % para el parámetro μ de la distribución.
 - Así tenemos una probabilidad del 95% de encontrar μ en ese intervalo.
- Recordemos que al tomar una muestra de tamaño n de una variable con media μ y varianza σ^2 y distribución cualquiera (en particular una normal también), la distribución muestral de la media verifica que $E[\bar{x}] = \mu$ y $\text{Var}[\bar{x}] = \sigma^2/n$.
- En este caso la distribución muestral de la media muestral \bar{x} sigue una normal $N(\mu, \sigma_0/(n)^{0.5})$ ([transparencia 226](#)).
- Esto lo podemos aprovechar para construir el siguiente estadístico como variable aleatoria z que sigue una $N(0,1)$, que viene determinado por
 - $z = (\bar{x} - \mu)/(\sigma_0/\sqrt{n})$

Estimación por intervalos: un ejemplo simple (para entender el problema)

- Así la nueva variable se distribuye según una $N(0,1)$
- Por favor notar que es importante que el nuevo estadístico tenga una distribución conocida e independiente del parámetro que queremos estimar en este caso que es μ a través de la media muestral \bar{x} :
 - Ya que esto nos permite dada la nueva variable z construir una expresión del tipo, sabiendo que z es una normal:
 - $P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) =$
 $P(-z_{\alpha/2} \leq (\bar{x} - \mu)/(\sigma_0/\sqrt{n}) \leq z_{\alpha/2}) = 1 - \alpha$
 - Donde $z_{\alpha/2}$ es un valor de la normal estándar tal que cumple:
 - $P(z > z_{\alpha/2}) = 1 - \Phi(z_{\alpha/2}) = \alpha/2$, siendo Φ la función distribución normal.
- Esto nos permite estimar el parámetro μ a través de un intervalo de confianza de la siguiente manera.

Estimación por intervalos: un ejemplo simple (para entender el problema)

- Así la expresión anterior nos permite determinar, de manera independiente de μ , el valor $z_{\alpha/2}$ que delimita una probabilidad $(1-\alpha)$ dentro del intervalo centrado en cero $(-z_{\alpha/2}; z_{\alpha/2})$.
- En este caso, para la distribución $N(0, 1)$, y $(1-\alpha)=95\%$ y el valor $|z_{\alpha/2}|$ es aproximadamente 1,96 (mirar las tablas del final del libro).



Estimación por intervalos: un ejemplo simple (para entender el problema)

- Por lo tanto podemos poner para $\alpha=5\%$ según la figura anterior de la normal, teniendo en cuenta que 1.96 es el valor aproximado del punto percentil 97.5 de la distribución normal:
 - $P(-1.96 \leq (\bar{x} - \mu)/(\sigma_0/\sqrt{n}) \leq 1.96) = 0.95$
 - Tabla libro, pag 618, el área mas cercana es 0,97500 (que es el valor exacto que buscamos) que corresponde al percentil 1,96 de la normal.
- Así si despejamos μ obtenemos el intervalo de confianza siguiente:
 - $P(\bar{x} - 1.96 \frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 \frac{\sigma_0}{\sqrt{n}}) = 0.95$

Estimación por intervalos: un ejemplo simple (para entender el problema)

- Así la expresión $P(\bar{x} - 1.96 \frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 \frac{\sigma_0}{\sqrt{n}}) = 0.95$ lo que quiere decir básicamente es que:
- Para la **estimación de la media** con una **muestra de tamaño n** en una **población normal** de **desviación** típica conocida e igual a σ_0 .
 - **Podemos decir** que antes de observar la muestra y calcular el estimador, \bar{x} , se pueden predecir las **discrepancias** esperadas entre el estimador (\bar{x}) y el parámetro (μ) que vienen dadas por el siguiente intervalo de confianza el 95% de las veces que se observa la muestra:
 - $|\bar{x} - \mu| \leq 1,96 \frac{\sigma_0}{\sqrt{n}}$

Estimación por intervalos: método del pivote

- El método anterior se puede generalizar dando origen al conocido **método del pivote** para la construcción de intervalos de confianza.
- Así este se basa en la **elección de una variable aleatoria** que sea **función de la muestra** y del **parámetro a estimar**, con una serie de condiciones para la función:
 - Que sea una **función continua y monótona** del parámetro.
 - Que su **distribución sea conocida e independiente del parámetro a estimar**.

Estimación por intervalos: método del pivote

- Llamemos $g(\vartheta, X)$ a la variable escogida y que recibe el nombre de **estadístico pivote**.
- Vemos que esta variable depende del parámetro a estimar, ϑ , y la muestra escogida, X .
- Bajo estas condiciones, fijado el nivel de confianza $(1-\alpha)100\%$, siempre es posible encontrar los valores a y b tales que $P(a \leq g(\vartheta, X) \leq b) = (1-\alpha)$, siempre que la distribución sea conocida para la probabilidad, P , anterior.
- Por las condiciones exigidas sobre el estadístico, será posible despejar ϑ y por tanto obtener los límites para el intervalo.

Estimación por intervalos: método del pivote

- Recordemos que las condiciones que se exigen para el estadístico pivote son:
 1. Que sea una función continua y monótona del parámetro.
 2. Que su distribución sea conocida e independiente del parámetro a estimar.
- Debido a la primera condición podemos despejar el estimador para deducir el intervalo de confianza:
 - $P(g^{-1}(a, X) \leq \vartheta \leq g^{-1}(b, X)) = (1 - \alpha)$
- Siendo $\vartheta_1 = g^{-1}(a, X)$ y $\vartheta_2 = g^{-1}(b, X)$ los límites del intervalo deseado, para el estimador ϑ .

Estimación por intervalos: método del pivote

- Debido a la primera condición podemos asegurar que existen $g^{-1}(a, X)$ y $g^{-1}(b, X)$.
- Se puede demostrar que si una función f es continua y monótona en un intervalo $[a, b]$, entonces existe la inversa en el intervalo $[f(a), f(b)]$, y es también monótona y continua.
- Una función monótona es aquella que o bien es creciente o bien es decreciente.
- Debido a la segunda condición podemos calcular la confianza a través de la función de probabilidad (ya que la distribución es conocida):
 - $P(g^{-1}(a, X) \leq \vartheta \leq g^{-1}(b, X)) = (1 - \alpha)$

Estimación por intervalos: método del pivote

- Dado $P(a \leq g(\vartheta, X) \leq b) = (1 - \alpha)$, notar que **los valores a y b que la verifican en general no son únicos**.
- La elección se hace generalmente buscando que el **intervalo tenga la máxima precisión**, es decir, la longitud mínima.
- Esta elección depende de la distribución del estadístico pivote.
- Para distribuciones simétricas y unimodales (distribución Normal o t de Student, por ejemplo) se consigue tomando el intervalo centrado, es decir, dejando una probabilidad de $\alpha/2$ a cada lado, como hemos visto en la figura anterior.

Estimación por intervalos: principales estadísticos pivote

➤ Intervalos para medias de poblaciones normales (varianza conocida de la población):

- Para poblaciones normales que se conoce la varianza, σ , ya sabemos cual es el estadístico pivote para el cálculo de intervalos de confianza para la media μ :

- $z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$, en este caso z se distribuye como una $N(0,1)$.

- $P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) =$

$$P(-z_{\alpha/2} \leq (\bar{x} - \mu)/(\sigma/\sqrt{n}) \leq z_{\alpha/2}) = 1 - \alpha$$

- Donde $z_{\alpha/2}$ es un valor de la normal estándar tal que:

- $P(z > z_{\alpha/2}) = 1 - \Phi(z_{\alpha/2}) = \alpha/2$, siendo Φ la función distribución normal.

- Así los intervalos de confianza para μ a un nivel de confianza $(1 - \alpha)$ serán :

- $\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Estimación por intervalos: principales estadísticos pivote

➤ Intervalos para medias de poblaciones normales (varianza conocida de la población):

- Recordar que los valores que se han escogido son simétricos para generar el intervalo más corto posible.
- Para la distribución de confianza despejamos de $z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$,
- $\mu = \bar{x} + z(\sigma/\sqrt{n})$, así al variar α , como \bar{x} es constante, la única variable aleatoria es z .
- Así, la distribución generada es normal, con media \bar{x} y varianza σ/\sqrt{n} .
- Esta distribución resume la incertidumbre existente respecto al valor desconocido μ .

Estimación por intervalos: principales estadísticos pivote

- Cual es el mínimo tamaño muestra necesario para obtener una precisión dada para estimar el la media en un intervalo de confianza dado. **Vamos a hacer un ejemplo.**
- Supongamos que la altura de los individuos de cierta población sigue una distribución $N(\mu, 7,5)$ estando las unidades en cm. Hallar el mínimo tamaño muestral necesario para estimar la altura media con un error inferior a 2 cm y con una confianza del 90%.
- Lo que queremos es que $|\bar{x} - \mu| < 2$ y como queremos una confianza del 90% entonces $z_{\alpha/2} = z_{0,05} = 1,645$ (tabla libro, pag 618, el área mas cercana es 0.95053 que corresponde a 1,65).
- Así sabemos que el intervalo de confianza de para la media de la altura viene determinado por: $\bar{x} - 1,645 \frac{7,5}{\sqrt{n}} \leq \mu \leq \bar{x} + 1,645 \frac{7,5}{\sqrt{n}}$ (el valor de 1,645 está calculado de manera más exacta).
- Una expresión equivalente es: $1,645 \frac{7,5}{\sqrt{n}} \leq \bar{x} - \mu \leq 1,645 \frac{7,5}{\sqrt{n}}$
- O lo que es lo mismo: $|\bar{x} - \mu| \leq 1,645 \frac{7,5}{\sqrt{n}}$
- Sustituyendo por su valor $2 \leq 1,645 \frac{7,5}{\sqrt{n}}$, y despejando n obtenemos $n > 38,05$
- Es decir con una muestra de $n=39$, el 90% de las veces el error en la estimación de la media de la altura para la población a través de la muestra es menor que 2 cm.

Estimación por intervalos: principales estadísticos pivote

➤ Intervalos para medias de poblaciones normales (varianza desconocida de la población):

- Para este caso se utiliza una variable de Student para estimar el intervalo de confianza de μ : $t = (\bar{x} - \mu)/(\hat{s}/\sqrt{n})$ con $n-1$ grados de libertad.
- Además, el estadístico obtenido no depende de σ , y es función monótona de μ (esto es justo lo que necesitamos para el método del pivote), por lo tanto podemos decir que:

- $P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) =$

$$P(-t_{\alpha/2} \leq (\bar{x} - \mu)/(\hat{s}/\sqrt{n}) \leq t_{\alpha/2}) = 1 - \alpha$$

- Donde $t_{\alpha/2}$ es un valor de una t de Student tal que cumple:
- $P(t > t_{\alpha/2}) = \alpha/2$, y por tanto el intervalo de confianza para μ a un nivel de confianza $(1 - \alpha)$ será:

- $\bar{x} - t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}}$

Estimación por intervalos: principales estadísticos pivote

- **Ejemplo:** En una empresa se anunció que los salarios el año pasado crecieron un promedio del 3,5%. Un grupo de trabajadoras toma una muestra de los incrementos que han recibido una muestra de 10 mujeres obteniendo los siguientes incrementos: 3%, 3%, 5%, 1%, 1%, 2%, 1%, 1,5%, 2%, 2%. Construir un intervalo de confianza para el incremento medio experimentado por la remuneración de las mujeres en esta empresa.
- La media de los incrementos es $\bar{x} = 2,36$ y la desviación típica de los 10 incrementos es $\hat{s} = 1,50$.
- Para un intervalo de confianza del 95% se requiere el percentil de la distribución t de Student con 9 grados de libertad que es 2,26 (tablas libro, pag 619).
- Por lo tanto el intervalo es $2,36 - 2,26 \frac{1,5}{\sqrt{10}} \leq \mu \leq 2,36 + 2,26 \frac{1,5}{\sqrt{10}}$
- Saliendo un intervalo de $1,29 \leq \mu \leq 3,44$, como este intervalo no incluye el 3,5% como valor posible, se puede concluir que existe una fuerte evidencia (al 95%) de que las mujeres han recibido un incremento salarial menor que la media de los trabajadores.

Estimación por intervalos: principales estadísticos pivote

- Intervalo para varianzas de poblaciones normales:
 - Para construir un intervalo de confianza para la estimación de la varianza de una población normal, tenemos en cuenta que
 - $\frac{ns^2}{\sigma^2} = \frac{(n-1)\hat{s}^2}{\sigma^2}$ se distribuye como χ_{n-1}^2 .
 - Por lo tanto, determinando dos valores χ_a^2 y χ_b^2 que dejen entre si $(1 - \alpha)$ de la distribución χ_{n-1}^2 ya lo tenemos:
 - $P(\chi_a^2 \leq \frac{ns^2}{\sigma^2} \leq \chi_b^2) = 1 - \alpha$, o expresado como $P(1/\chi_a^2 \geq \frac{\sigma^2}{ns^2} \geq 1/\chi_b^2) = 1 - \alpha$
 - Por tanto podemos encontrar el intervalo para estimar σ^2 a un nivel de confianza $(1 - \alpha)$ de la siguiente forma:
 - $ns^2/\chi_a^2 \geq \sigma^2 \geq ns^2/\chi_b^2$ (Ver ejercicio 8.2)
 - **Existen muchos más estadísticos pivote que pueden consultar en el capítulo 8 del libro.**

Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- La **media muestral** es un estimador con una propiedad muy especial:
 - **Su precisión** (inversa de la varianza de la distribución muestral) se **puede conocer a través de la varianza de la muestra**:
 - Recordar que para estimar la varianza de la distribución muestral de la media hacíamos $E[(\bar{x} - \mu)^2] = Var(\bar{x}) = \sigma^2/n$.
 - Así si estimamos la varianza σ^2 a través de la varianza de la muestra $\hat{\sigma}^2$ entonces ponemos que $\hat{\sigma}^2(\bar{x}) = \frac{\hat{\sigma}^2}{n} = \frac{1}{n(n-1)} \sum (x_i - \bar{x})^2$
 - La inversa de este valor, $1/\hat{\sigma}^2(\bar{x})$, es la precisión del estimador media muestral.
 - Esta expresión es válida en general, ya que no depende del modelo de distribución de probabilidad que genera la muestra, solo depende de los datos muestrales
 - Ser conscientes que al estimar cualquier otro parámetro o característica de la población (como varianza, curtosis, asimetría, etc.), la precisión de la estimación depende de la distribución que genera los datos.

Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- Si utilizamos un estimador de máxima verosimilitud podemos conocer su varianza asintótica, pero esa medida puede ser poco precisa para muestras pequeñas.
- Existen métodos de simulación por computador como estimación **jackknife** y **bootstrap** que son generales para obtener precisión de un estimador de forma aproximada sin hacer hipótesis respecto a las distribución.
- Estos métodos hoy en día se pueden usar por la potencia y rapidez de los ordenadores digitales.
- **Jackknife:** Quenouille, M. H. (September 1949). "Problems in Plane Sampling". *The Annals of Mathematical Statistics*. **20** (3): 355–375. [doi:10.1214/aoms/1177729989](https://doi.org/10.1214/aoms/1177729989). [JSTOR 2236533](https://www.jstor.org/stable/2236533).
- **Bootstrap:** [Efron, B. \(1979\). "Bootstrap methods: Another look at the jackknife". *The Annals of Statistics*. **7** \(1\): 1–26.](#)

Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- **Bootstrap se basa en calcular directamente la varianza del estimador considerando la muestra como si fuese toda la población y aplicando el método de Montecarlo (sección 5.7.2 del libro si queréis saber más sobre montecarlo) para obtener réplicas de la muestra.**

Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- Siempre se parte de una **muestra de datos** (x_1, \dots, x_n) y se realiza el siguiente procedimiento:
- **Primero:** Consideramos la muestra como una población de una variable que toma los n valores (x_1, \dots, x_n) con probabilidad $1/n$. De esta muestra se extrae una m.a.s. de tamaño n (igual que el tamaño de muestra) mediante el método de Montecarlo. Es decir obtenemos una muestra al azar con reemplazamiento de los valores observados. Notar que esta muestra generada no coincidirá, en general, con la muestra original. Supongamos que la primera muestra obtenida es (y_1^1, \dots, y_n^1) .

Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- **Segundo:** Calculamos en la muestra generada anterior el estimador cuya precisión queremos estimar:
 - $\hat{\vartheta}_1 = \hat{\vartheta}(y_1^1, \dots, y_n^1)$
- **Tercero:** Repetimos los pasos **primero** y **segundo** un número B de pasos grande (por ejemplo, 1000 veces), así obtenemos una secuencia del estimador $\hat{\vartheta}_1, \dots, \hat{\vartheta}_B$ en cada uno de las B m.a.s. Así la estimación de su media y su varianza es:
 - $\hat{\vartheta}_m = 1/B \sum \hat{\vartheta}_i, \quad \text{Var}(\hat{\vartheta}) = 1/B \sum (\hat{\vartheta}_i - \hat{\vartheta}_m)^2, \quad \forall i \in (1, \dots, B)$

Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- Se puede demostrar que en condiciones generales, este método obtiene asintóticamente la varianza del estimador $\hat{\vartheta}$, y que el intervalo de confianza a un nivel $1 - \alpha$ se puede obtener de la distribución de B valores de $\hat{\vartheta}_i$.
- Para el intervalo de confianza se obtienen los valores $\hat{\vartheta}_{\text{INF}}$ y $\hat{\vartheta}_{\text{SUP}}$ tales que:
 - $P(\hat{\vartheta}_{\text{INF}} \leq \hat{\vartheta}_i \leq \hat{\vartheta}_{\text{SUP}}) = 1 - \alpha$
- Los valores del intervalo de confianza $\hat{\vartheta}_{\text{INF}}$ y $\hat{\vartheta}_{\text{SUP}}$ se calculan ordenando los $\hat{\vartheta}_i$ y tomando los valores situados en las posiciones $[B \times \alpha/2]$ y $[B \times (1 - \alpha/2)]$, donde el símbolo $[]$ significa redondear al entero más cercano.
- Darse cuenta que suponemos que los B valores cubren el 100% de la distribución de la varianza del estimador $\hat{\vartheta}$.
- Ver el ejemplo 8.6 y 8.7, que se puede simular en Python también.

Inferencia Estadística: Estimación por intervalos

Estimación bayesiana

- Introducción
- Distribuciones a priori
 - Distribución conjugada a priori
- Estimación puntual
- Estimación de poblaciones normales con varianza conocida

Estimación bayesiana: introducción

- Los **métodos de estimación que hemos visto hasta ahora** funcionan **bien** con **muestras grandes**.
- Veamos varios ejemplos con muestras pequeñas:
- **Ejemplo 1:** Supongamos que nos preguntamos por la proporción de estudiantes que leen un determinada novela, y que en una muestra pequeña (30 estudiantes por ejemplo) obtenemos 0 personas que han leído el libro.
- Qué inferencia podemos hacer sobre el parámetro p , suponiendo un modelo binomial.
- Según lo que hemos visto, uno puede calcular el estimador máximo-verosímil del parámetro de una población binomial.

Estimación bayesiana: introducción

- En este caso si uno hace los cálculos aprendidos (función de verosimilitud de p , y luego optimización del parámetro derivando respecto de p e igualando a cero) la **estimación máximo verosímil del parámetro p es la frecuencia en la muestra**, que, en este caso, y nos ha salido cero.
- Así la estimación máximo verosímil del número de estudiantes que han leído esta novela es cero y para cuantificar la precisión de esta estimación nos encontramos que la varianza estimada es cero, o la **precisión infinita**.
- Así no parece razonable aplicar lo métodos estudiados.

Estimación bayesiana: introducción

- El problema es que estamos aplicando un método máximo verosímil, que tiene buenas propiedades en muestras grandes, pero no en pequeñas o medianas muestras.
- **Ejemplo2:** Como segundo ejemplo, supongamos que tratamos de estimar la edad del más veterano de los estudiantes de una universidad mediante esa misma muestra de tamaño 30.
- Como desconocemos la distribución de edades, podríamos estimar como valor máximo de la variable edad el mayor valor observado en la muestra.

Estimación bayesiana: introducción

- Supongamos que la persona de mayor edad es de 21 años.
- Esta estimación es intuitivamente muy deficiente.
- Es difícil que el más veterano este en la muestra.
- **Ejemplo3:** Como tercer ejemplo supongamos que una moneda de un euro, la tiramos 10 veces y obtenemos 7 caras y 3 cruces. Con esto queremos decir que la probabilidad de cara en esta moneda es 0,7. Claramente no.

Estimación bayesiana: introducción

- Estos tres ejemplos tienen en común la existencia de cierta información a priori respecto al parámetro que tratamos de estimar, que no se tiene en cuenta en el proceso de inferencia: **información a priori.**
- Ignorar la información inicial que tenemos respecto a un parámetro a estimar no es importante si la muestra es grande, pero puede serlo cuando la muestra es pequeña.
- Es decir si la muestra es muy grande podemos despreciar al información a priori frente a la gran cantidad de información que tenemos en la muestra.

Estimación bayesiana: introducción

- Sin embargo con una pequeña muestra desperdiciar la información a priori no nos permite hacer un inferencia correcta, ya que la información a priori es significativa frente a los datos (la muestra).
- **Por tanto la inferencia bayesiana es un procedimiento general para combinar nuestra información a priori con la muestra para obtener una inferencia que tenga en cuenta toda la información existente en el problema.**

Estimación bayesiana: introducción

- En el **enfoque bayesiano** un **parámetro no es una constante** desconocida, **sino una variable aleatoria** sobre la que podemos establecer a priori una distribución de probabilidad que refleje nuestro conocimiento del problema a priori previo a realizar el proceso de inferencia.
- El proceso de inferencia se obtiene aplicando el cálculo el **teorema de Bayes**: combina la **información a priori** con la **información de la muestra** y se obtiene la **distribución del parámetro condicionada a la información disponible**.
- Por tanto, suponemos que **antes de tomar la muestra** se dispone de cierta información respecto al parámetro que se representa mediante una **distribución inicial o a priori**, $p(\theta)$.

Estimación bayesiana: introducción

- Supongamos que $f(X|\theta)$ es la **probabilidad de obtener una muestra para cada valor posible del parámetro**.
- Después de tomar la muestra $X = (x_1, \dots, x_n)$, en la función $f(X|\theta)$ los datos son fijos, ya que han sido observados, y los parámetros θ son variables para esa muestra observada.
- Así **cuando la muestra se observa** $f(X|\theta) = \ell(\theta|X)$, que es lo que habíamos dicho que era la función de **verosimilitud**.

Estimación bayesiana: introducción

- Podemos combinar las cantidades anteriores a través del teorema de Bayes para encontrar la probabilidad a posteriori $p(\theta|X)$:

- $$p(\theta|X) = \frac{f(X|\theta)p(\theta)}{\int f(X|\theta)p(\theta)d\theta}$$

- La distribución a posteriori $p(\theta|X)$ contiene toda la información para hacer inferencias respecto al parámetro.
 - Si se desea un estimador puntual, se tomará la media o la moda de dicha distribución
 - Si se desea un intervalo de confianza, se tomará la zona que encierre una probabilidad fijada en dicha distribución.
- En resumen, una vez obtenida la distribución de probabilidad del parámetro, el problema de estimación queda resuelto de manera precisa.

Estimación bayesiana: introducción

- Observar que el denominador $m(X) = \int f(X|\theta)p(\theta)d\theta$ y como función de X (datos muestrales) representa la distribución marginal de los datos, con independencia de los valores de los parámetros (esta integrado sobre el parámetro).
- Esta distribución se suele denomina **distribución predictiva** y observando la expresión nos damos cuenta que es una media ponderada de las verosimilitudes $f(X|\theta)$ por las probabilidades que la distribución a priori asignada a los posibles valores del parámetro θ a estimar.

Estimación bayesiana: introducción

- Cuando observamos la muestra este denominador es una constante, y el cálculo de la probabilidad a posteriori $p(\theta|X)$ se simplifica.
- Por lo tanto el denominador es solo es una constante de normalización para que la integral de numerador sea la unidad:
 - De esta manera la densidad de probabilidad $p(\theta|X)$, esta bien definida.
- Así podemos poner que
 - $p(\theta|X) = kf(X|\theta)p(\theta) = k\ell(\theta|X)p(\theta)$.

Estimación bayesiana: introducción

- Así tenemos $p(\theta|X) = k\ell(\theta|X)p(\theta)$, que nos dice que podemos calcular la distribución posteriori multiplicando, para cada valor del parámetro θ , su verosimilitud $\ell(\theta|X)$ por su probabilidad a priori $p(\theta)$.
- Observar que la constante k es irrelevante para la forma de la probabilidad a posteriori $p(\theta|X)$, como hemos dicho es un factor de normalización que se puede determinar al final con la condición de que la integral de $p(\theta|X)$ sea la unidad.
- Así el teorema de Bayes se puede resumir como:
 - $A \text{ posteriori} \propto \text{verosimilitud} \times A \text{ priori}$
- Donde el símbolo \propto significa proporcional.

Estimación bayesiana: introducción

- En el caso particular de que $p(\theta)$ sea aproximadamente constante sobre el rango de valores en los que la verosimilitud no es nula, se dice que la distribución a priori, $p(\theta)$, es no informativa, y por tanto la distribución a posteriori $p(\theta|X)$, vendrá determinada únicamente por la función de verosimilitud $\ell(\theta|X)$.
- Una ventaja adicional del enfoque bayesiano es su facilidad para procesar información secuencialmente. Esto es muy importante para afinar la estimación.
- Supongamos que después de calcular la distribución a posteriori con la observación de una muestra X observamos una nueva muestra de la misma población Y , independiente de la primera.
- Entonces, la nueva distribución a posteriori final se puede actualizar con la distribución inicial a posteriori a través de la expresión:
 - $p(\theta|XY) = k\ell(\theta|Y)p(\theta|X)$.

Estimación bayesiana: introducción

- Naturalmente la expresión anterior se puede obtener de la aplicación de la estimación bayesiana a una muestra ampliada $[X, Y]$ donde la muestra X e Y son independientes:
 - $p(\theta|XY) = k\ell(\theta|XY)p(\theta) = kf(XY|\theta)p(\theta) = kf(X|\theta)f(Y|\theta)p(\theta) = k\ell(\theta|X)\ell(\theta|Y)p(\theta) = k\ell(\theta|Y)\ell(\theta|X)p(\theta) = k\ell(\theta|Y)p(\theta|X) = p(\theta|XY)$
- Así la estimación bayesiana proporciona pues un procedimiento automático para expresar el aumento de nuestro conocimiento respecto al parámetro a medida que se recibe información adicional.

Estimación bayesiana: distribuciones a priori

- La mayor dificultad práctica del enfoque bayesiano es cómo especificar la distribución a priori: normalmente la información de que disponemos es cualitativa y el enfoque bayesiano requiere que establezcamos una distribución de probabilidad sobre sus valores.
- Así se pueden considerar cuatro casos diferentes.
- **Primero:** La distribución a priori proviene de estudios anteriores y se conoce objetivamente. Por ejemplo, supongamos que tratamos de determinar
- **Segundo:** La distribución a priori puede ser importante respecto a la muestral, pero la información existente es subjetiva y no formalizada. Por ejemplo podemos elegir una distribución a priori que refleje globalmente nuestra opinión sobre la distribución a priori del parámetro, en particular la moda a priori y el rango de valores posibles, si la distribución es o no simétrica, etc.

Estimación bayesiana: distribuciones a priori

- **Tercero:** La información a priori es pequeña con relación a la información muestral. Podemos elegir una distribución a priori que refleje globalmente nuestra opinión, en particular la moda a priori y el rango de valores posibles, pero sin preocuparnos mucho del resto de los detalles. En este caso se suelen escoger distribuciones conjugadas (lo vemos después).
- **Cuarto:** La información a priori es despreciable frente a la información muestral, o no queremos tenerla en cuenta en el proceso de inferencia:
 - En este caso podemos utilizar los métodos clásicos que hemos estudiado anteriormente.
 - O también podemos utilizar el enfoque bayesiano con una distribución a priori *no informativa* o de *referencia*.

Estimación bayesiana: distribución conjugada

- La idea es expresar aproximadamente nuestra información a priori con una distribución que facilite el análisis.
- Un ejemplo es utilizar una familia de distribuciones a priori que tiene la misma forma que la verosimilitud, de manera que la posterior pueda calcularse fácilmente al pertenecer a la misma familia que la priori.
- A estas familias se las denomina *conjugadas*.
- Una clase \mathcal{C} de distribuciones a priori para un parámetro θ es conjugada si cuando la a priori pertenece a esa clase, $p(\theta) \in \mathcal{C}$ entonces también lo hace la a posteriori $p(\theta|X) \in \mathcal{C}$.
- La distribución conjugada a priori se elige tomando como distribución la verosimilitud, y modificando los valores de las constantes para que la función resultante sea una función de densidad y tenga características coincidentes con nuestra información a priori.

Estimación bayesiana: distribución conjugada

- Por ejemplo, supongamos que queremos hacer estimar el parámetro en una modelo normal de varianza conocida.
- La verosimilitud es
 - $\ell(\theta) = k \exp\left(-\frac{n}{2\sigma^2}(\bar{x} - \theta)^2\right)$
- Y la a priori conjugada la escogemos como
 - $p(\theta) = k \exp\left(-\frac{n_0}{2\sigma^2}(\theta - \mu_0)^2\right)$
- Siendo μ_0 un parámetro para variar la media de la distribución y n_0 otro para ajustar la varianza, que determinan la forma de la distribución $p(\theta)$, para que la función resultante coincida con nuestra opinión sobre la a priori.

Estimación bayesiana: estimación puntual

- Si queremos hacer estimación puntual con la a posteriori es necesario elegir un valor único para el parámetro.
- Para ello podemos optar por las diferentes posibilidades:
- **Primera:** Seleccionar el máximo (la moda) de la distribución a posteriori, que es el valor más probable. Cuando la información que nos da la a priori sea pequeña con relación a la proporcionada por la verosimilitud, la a posteriori será análoga a la verosimilitud y su moda es el estadístico máximo-verosímil. Por tanto, en este caso el enfoque bayesiano coincide con el MV.
- **Segunda:** Definir un criterio de optimalidad y deducir el estimador a partir de él. Esto equivale a definir *una función de pérdida*, $g(\vartheta, \hat{\vartheta})$ que indique la penalización de tomar $\hat{\vartheta}$ como estimador cuando el valor real es ϑ .

Estimación bayesiana: estimación puntual

- La función de pérdida más frecuente escogida es la cuadrática $g(\boldsymbol{\vartheta}, \hat{\boldsymbol{\vartheta}}) = k(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})^2$.
- El criterio de elección será escoger como estimador aquel valor $\hat{\boldsymbol{\vartheta}}$ que haga en promedio la pérdida mínima.
- Esto quiere decir $E[g(\boldsymbol{\vartheta}, \hat{\boldsymbol{\vartheta}})] = kE[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})^2]$, donde la esperanza se toma respecto a la distribución de $\boldsymbol{\vartheta}$.
- A la pérdida promedio se denomina *riesgo del estimador*.

Estimación bayesiana: estimación de poblaciones normales con varianza conocida

- Supongamos que se desea estimar la media de una población normal con varianza conocida y que la información inicial respecto a μ se traduce en una distribución a priori $N(\mu_0, \sigma_0)$, suponiendo una distribución conjugada del tipo que hemos visto anteriormente para la distribución a priori:
 - $f(\mu) = k \exp\left(-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right)$
- Así la verosimilitud para una normal de varianza σ^2
 - $\ell(\mu|X) = k' \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right)$

Estimación bayesiana: estimación de poblaciones normales con varianza conocida

➤ Y la a posteriori conjugada es entonces:

➤
$$f(\mu|X) = k'' \ell(\mu|X) f(\mu) = k' \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right) k \exp\left(-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right)$$



Estimación bayesiana: estimación de poblaciones normales con varianza conocida

- Su exponente puede escribirse de la siguiente forma (ejercicio 9.6):

$$\begin{aligned} -\frac{n}{2\sigma^2}(\bar{x} - \mu)^2 - \frac{1}{2\sigma_0^2}(\mu - \mu_0)^2 &= \frac{n}{2\sigma^2}(\bar{x} - \mu)^2 + \frac{1}{\sigma_0^2}(\mu - \mu_0)^2 \\ &= \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)(\mu - \mu_p)^2 + \left(\frac{n}{n\sigma_0^2 + \sigma^2}\right)(\bar{x} - \mu_0)^2 \end{aligned}$$

- Con $\mu_p = \frac{\frac{n}{\sigma^2}\bar{x} + \frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$

- Así al final tenemos que la expresión para el producto de la verosimilitud ($\ell(\mu|\bar{x}) = f(\bar{x}|\mu)$) por la a priori viene determinada por: $f(\bar{x}|\mu)f(\mu) = \exp\left(\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)(\mu - \mu_p)^2 + \left(\frac{n}{n\sigma_0^2 + \sigma^2}\right)(\bar{x} - \mu_0)^2\right)$

Estimación bayesiana: estimación de poblaciones normales con varianza conocida

- Por otro lado sabemos por el Teorema de Bayes que $f(\bar{x}|\mu)f(\mu) = f(\mu|\bar{x})f(\bar{x})$
- Así este segundo miembro está compuesto por la distribución a posteriori ($f(\mu|\bar{x})$) y la distribución **predictiva** ($f(\bar{x})$), que no depende de \bar{x} , es decir:
- $f(\mu|\bar{x}) = \exp\left(-\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)(\mu - \mu_p)^2\right)$ y $f(\bar{x}) = \exp\left(-\left(\frac{n}{n\sigma_0^2 + \sigma^2}\right)(\bar{x} - \mu_0)^2\right)$
- Supongamos que definimos por $p_{\bar{x}}$ y p_0 a las precisiones muestral y a priori, entonces la distribución a posteriori es una normal con parámetros:
$$\mu_p = \frac{\frac{n}{\sigma^2}\bar{x} + \frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} = \frac{p_{\bar{x}}\bar{x} + p_0\mu_0}{p_{\bar{x}} + p_0} \text{ y } \frac{1}{\sigma_p^2} = p_p = p_{\bar{x}} + p_0$$
- **Resumiendo la media de la a posteriori es una combinación lineal de la a priori y la muestral, con pesos que dependen de la precisión relativa. La precisión final es la suma de la inicial (p_0) y la verosimilitud ($p_{\bar{x}}$).**

Bibliografía y lecturas relacionadas:

- Fundamentos de estadística. Daniel Peña Sánchez Ribera. Alianza Editorial, 2001 o 2008.
- Data analysis Statistical and computational methods for scientists and engineers Recurso electrónico 4th.ed. Autor Brandt, Siegmund.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer Second Edition February 2009.
- Probabilidad y Estadística con Python.
- Entre percentiles, cuartiles y cuantiles. Posted on 21 febrero, 2015 by Jesús Soto.
- Cuartiles.
- *NIST/SEMATECH e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/>, 09/22/2015.
- <http://www.itl.nist.gov/div898/handbook/eda/section3/normprpl.htm>
- Workshop on Big Data and Statistics UDC, junio 2015
- <http://bigdata.sgapeio.es/index.php/en/sesion-i>

Muchas Gracias!!!!

¿Peguntas?

Breve Biografía del Profesor:

Francisco de Borja Rodríguez Ortiz received his degree in Applied Physics in 1992. Then he worked at the Instituto de ingeniería del conocimiento (Universidad Autónoma de Madrid) until 1995. He received his PhD in Computer Science in 1999 from Universidad Autónoma de Madrid. Then he worked at the Nijmegen University in Holland, at the Institute for Nonlinear Science in University of California San Diego, at Centro de Neurociencias Integradas (CENI) in Santiago de Chile and at Instituto de Física de São Carlos in Universidad de São Paulo (IFSC, USP). Since 2002 he is “Profesor Titular” at the Escuela Politécnica Superior, Universidad Autónoma de Madrid. FBRO has a 21 year teaching experience at the Departamento de Ingeniería Informática, Universidad Autónoma de Madrid where he is a lecturer in Cryptography, Data Structure Information, Scientific Computing, Pattern Recognition, Basic Computer Science, Neurocomputation Fundamentals, Databases Systems and graduate courses in Computational Neuroscience, Biomedical Signal Processing and its and its applications, Biodevices. His areas of interest in investigation are Computational Neuroscience - Chemical Sensing - Artificial Neural Networks - Biologically inspired robotics - Information Theory and Clustering- Cryptographic Protocols in Anonymity.

Scientific publications of **Francisco de Borja Rodríguez Ortiz**:

<http://scholar.google.es/citations?user=7yEkpQIAAAAJ&hl=es>

Grupo de Neurocomputación Biológica, <http://www.ii.uam.es/~gnb>

Escuela Politécnica Superior, Universidad Autónoma de Madrid

f.rodriguez@uam.es

