Modelos matemáticos I

J. Héctor Morales Bárcenas

jhmb@xanum.uam.mx

Departamento de Matemáticas

Universidad Autónoma Metropolitana

Unidad Iztapalapa

Casa abierta al tiempo

Regresión lineal

Problema inverso lineal y discreto:

- Vector de datos d, N observaciones y un vector de parámetros x que deseamos determinar.
- Sistema lineal de ecuaciones $Ax = d$.
- Si $\mathrm{rank}(A) = N$, la matriz es de rango completo en sus columnas.
- Aún si así fuera, es posible que no halla solución exacta a la ecuación debido a que la dimensión del rango de A sea menor que N y que el vector de datos d contenga ruido y no esté dentro del rango de A.

Regresión lineal

- Una aproximación útil al problema es hallar el conjunto de valores de x tales que minimice en alguna medida el ajuste del modelo Ax y los datos d.

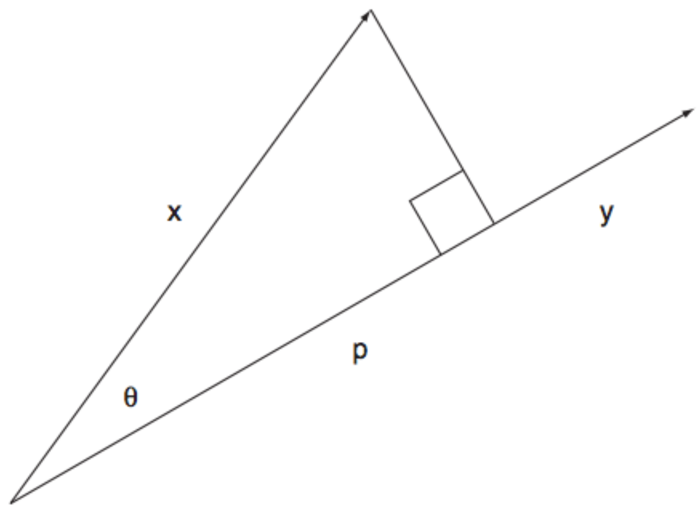
- Definimos al vector residuos r

$$r = d - Ax$$

- Una forma de medir o cuantificar su magnitud es por medio de la norma L_2

$$\text{mín}\,||Ax - d||_2$$

Regresión lineal



- Esto significa que en realidad el sistema es inconsistente y d no está en el espacio columan de A.
- La construcción de la solución aproximada x^+ consiste en proyectar a d en el rango de $\mathcal{R}(A)$

$$Ax^+ = \underset{\mathcal{R}(A)}{\mathrm{proj}}(d)$$

Regresión lineal

- Luego entonces, $Ax - d$ es perpendicular a $\mathcal{R}(A)$. En particular, cada columna de A es ortogonal a $Ax - d$. Así

$$A^T(Ax^+ - d) = 0,$$

o bien,

$$A^T A x^+ = A^T d.$$

- De donde obtenemos las ECUACIONES NORMALES

$$x^+ = (A^T A)^{-1} A^T d.$$

- Y, por lo tanto,

$$x_{L_2} = x^+.$$

Regresión lineal

- El problema de regresión lineal en el plano: determinar dos parámetros x_1 y x_2 de una línea, $y = x_1 + x_2 x$, que mejor ajuste a un conjunto de $N > 2$ datos.
- A partir del sistema de ecuaciones

$$Ax = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_N \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} = d.$$

- Aplicamos las ecuaciones normales:

$$x_{L_2} = (A^T A)^{-1} A^T d = \left(\begin{bmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_N \end{bmatrix} \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_N \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_N \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

Regresión lineal

- De donde se obtiene:

$$x_{L_2} = \begin{bmatrix} N & \sum_{i=1}^{N} a_i \\ \sum_{i=1}^{N} a_i & \sum_{i=1}^{N} a_i^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{N} d_i \\ \sum_{i=1}^{N} a_i d_i \end{bmatrix}$$

$$= \frac{1}{N\sum_{i=1}^{N} a_i^2 - \left(\sum_{i=1}^{N} a_i\right)^2} \begin{bmatrix} -\sum_{i=1}^{N} a_i^2 & -\sum_{i=1}^{N} a_i \\ \sum_{i=1}^{N} a_i & -N \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{N} d_i \\ \sum_{i=1}^{N} a_i d_i \end{bmatrix}.$$

Fórmula del error

- Supongamos que tenemos un conjunto de datos $i \in [1, N]$ dados como parejas (X_i, d_i) que siguen una tendencia lineal.
- Nuestro objetivo es determinar los parámetros a_1 y a_2 del modelo

$$y_M = a_1 x + a_2 \, .$$

- Queremos minimizar el error de mínimos cuadrados

$$E^2 = \frac{1}{N} \sum_{i=1}^{N} \left(d_i - y_M(X_i)\right)^2 = \frac{1}{N} \sum_{i=1}^{N} \left(d_i - a_1 X_i - a_2\right)^2 \, .$$

Fórmula del error

- Este error se puede escribir distribuyendo la suma

$$
\begin{aligned}
E^2(a_1, a_2) =& \frac{1}{N}\sum_{i=1}^{N} d_i^2 - 2a_1\frac{1}{N}\sum_{i=1}^{N} X_i d_i + a_1^2\frac{1}{N}\sum_{i=1}^{N} X_i^2 - 2a_2\frac{1}{N}\sum_{i=1}^{N} d_i \\
&+2a_1a_2\frac{1}{N}\sum_{i=1}^{N} X_i + a_2^2 \\
=& \langle d^2\rangle - 2a_1\langle X d\rangle + a_1^2\langle X^2\rangle - 2a_2\langle d\rangle + 2a_1a_2\langle X\rangle + a_2^2 \\
=& \langle \tilde{d}^2\rangle - 2a_1\langle \tilde{X}\tilde{d}\rangle + \left(a_2 - \langle d\rangle + a_1\langle X\rangle\right)^2 .
\end{aligned}
$$

Fórmula del error: puntos críticos

- La idea de optimizar este error (minimizarlo) implica hallar los puntos críticos en donde E^2 se anula:

$$\begin{aligned}\frac{\partial E^2}{\partial a_1} &= -2\langle \tilde{X}\tilde{d}\rangle + 2a_1\langle \tilde{X}^2\rangle + 2\langle X\rangle\left(a_2 - \langle d\rangle + a_1\langle X\rangle\right)\\ \frac{\partial E^2}{\partial a_2} &= 2\left(a_2\langle d\rangle + a_1\langle X\rangle\right).\end{aligned}$$

$$\begin{aligned}a_1 &= \frac{\langle \tilde{X}\tilde{d}\rangle}{\langle \tilde{X}^2\rangle}\\ a_2 &= \langle d\rangle - a_1\langle X\rangle.\end{aligned}$$

Fórmula del error

Finalmente, el modelo lineal dado los parámetros a y b:

$$y_M = \langle Y \rangle + \frac{\langle \tilde{X} \tilde{d} \rangle}{\langle \tilde{X}^2 \rangle} (x - \langle X \rangle)$$

El valor del error mínimo E^2 se puede hallar evaluando los valores críticos de a y b:

$$E^2 = \langle \tilde{d^2} \rangle - 2 \frac{\langle \tilde{X} \tilde{d} \rangle}{\tilde{X}^2} \langle \tilde{X} \tilde{d} \rangle + \frac{\langle \tilde{X} \tilde{d} \rangle}{\tilde{X}^2} = \langle \tilde{Y}^2 \rangle - \frac{\langle \tilde{X} \tilde{d} \rangle}{\langle \tilde{X}^2 \rangle}$$

que no es otra cosa que:

$$E^2 = \langle \tilde{d^2} \rangle \left(1 - r_{Xd}\right).$$

Aspectos estadísticos de la regresión lineal (Aster, 2005)

- Cuando los datos provienen de mediciones siempre poseen incertidumbre o variaciones aleatorias.
- Deseamos hallar la solución al problema de ajuste desde un esquema estadístico.

La **estimación de máxima verosimilitud** tiene la siguiente perspectiva:

1. Dado un **conjunto de datos** observacionales,
2. de los que poseemos alguna información de sus **propiedades estadísticas**,
3. y también poseemos un **modelo matemático** del **problema directo**,

Aspectos estadísticos de la regresión lineal

¿Cuáles son los parámetros del modelo que

reproducen mejor la tendencia de los datos?

- A las mediciones u observaciones se les asigna un función de densidad de probabilidad.

- El problema esencial es hallar los valores paramétricos más probables en términos del vector de datos d.

Aspectos estadísticos de la regresión lineal

- Supondremos que el conjunto de observaciones son independientes.
- Dado el vector de parámetros x, tenemos una densidad de probabilidad $f_i(d_i|x)$ para cada dato i.
- En general, estas densidades de probabilidad varian dependiendo de x.
- La **densidad de probabilidad conjunta** de un vector de observaciones independientes d es
$$f(d|x) = f_1(d_1|x)f_2(d_2|x)\cdots f_N(d_N|x)$$
- sólo podemos calcular la probabilidad de observar los datos en algún conjunto de valores dado los parámetros x, integrando $f(d|x)$ en tal conjunto.

Aspectos estadísticos de la regresión lineal

En la práctica medimos a un conjunto de datos y deseamos hallar "los mejores" valores paramétricos de un modelo en un sentido de **máxima verosimilitud**.

La función de **máxima verosimilitud**:

$$\begin{aligned} L(x|d) &= f(d|x) \\ &= f_1(d_1|x) f_2(d_2|x) \cdots f_N(d_N|x) \, . \end{aligned}$$

- De acuerdo con el principio de máxima verosimilitud, debemos seleccionar los valores paramétricos x de tal forma que se maximice dicha función.

- Los estimadores obtenidos mediante este método tienen propiedades estadísticas deseables.

Aspectos estadísticos de la regresión lineal

Supongamos que los datos tienen errores aleatorios independientes y normalmente distribuidos, con esperanza cero, y una desviación estándar de la i-ésima observación, d_i, igual a σ_i.

- La densidad de probabilidad de d_i es de la forma:

$$f_i(d_i|x) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(d_i - (Ax)_i)^2}{2\sigma_i^2}\right).$$

- La función de verosimilitud del conjunto completo de datos es el producto de las verosimilitudes individuales:

$$\begin{aligned} L(x|d) &= \Pi_{i=1}^N f_i(d_i|x) \\ &= \frac{1}{(2\pi)^{N/2}\Pi_{i=1}^N \sigma_i} \Pi_{i=1}^N \exp\left(-\frac{(d_i - (Ax)_i)^2}{2\sigma_i^2}\right). \end{aligned}$$

Aspectos estadísticos de la regresión lineal

El factor constante no afecta la maximización de L,

$$\max \Pi_{i=1}^{N} \exp\left(-\frac{(d_i - (Ax)_i)^2}{2\sigma_i^2}\right).$$

El logaritmo es una función monotonamente creciente, por lo que es equivalente resolver

$$\max \log\left(\Pi_{i=1}^{N} \exp\left(-\frac{(d_i - (Ax)_i)^2}{2\sigma_i^2}\right)\right) = \max\left\{-\sum_{i=1}^{N} \frac{(d_i - (Ax)_i)^2}{2\sigma_i^2}\right\}.$$

Aspectos estadísticos de la regresión lineal

Finalmente, si retornamos al problema de minimización cambiando el signo e ignorando el factor de $1/2$, el problema se convierte en

$$\min \sum_{i=1}^{N} \frac{(d_i - (Ax)_i)^2}{\sigma_i^2}$$

A parte de los factores $1/\sigma_i^2$ en cada término, este problema es el de mínimos cuadrados de $Ax = d$.

Para incorporar las desviaciones estándar en la solución, escalamos al sistema de ecuaciones. Sea W la matriz de pesos diagonal

$$W = \operatorname{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \cdots, \frac{1}{\sigma_N}\right).$$

Aspectos estadísticos de la regresión lineal

Sea entonces $A_w = W A$ y $d_w = W d$. Por lo que el sistema de ecuaciones con peso

$$A_w x = d_w.$$

Las ecuaciones normales se convierten, entonces,

$$x_{L_2} = \left(A_w^T A_w \right)^{-1} A_w^T d_w.$$

Por lo tanto, la solución al **problema de mínimos cuadrados** $A_w x = d_w$, es la solución de **máxima verosimilitud**.

Aspectos estadísticos de la regresión lineal

- La suma de los cuadrados de los residuos también nos da información estadística útil acerca de la calidad de los estimadores.

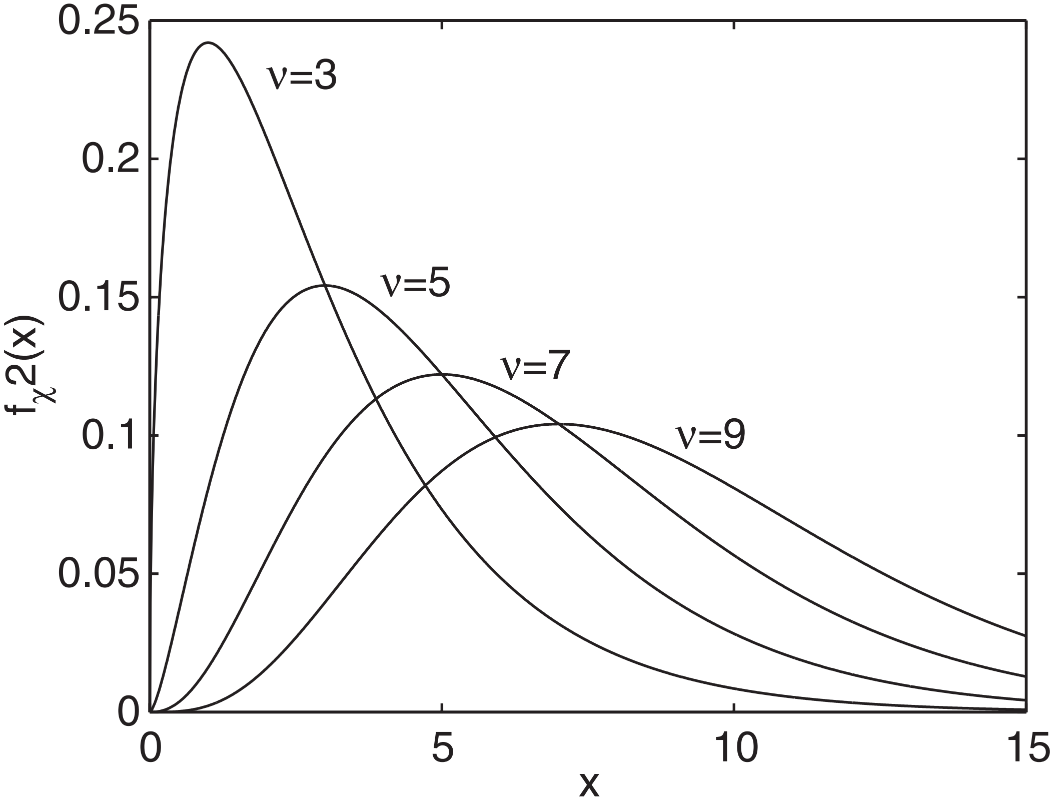
- La **estadística chi cuadrada** es

$$\chi^2_{\text{obs}} = \sum_{i=1}^{N} \frac{(d_i - (Ax)_i)^2}{\sigma_i^2}.$$

- Dado que χ^2_{obs} depende de los errores aleatorios de medición en d, por si misma es una variable aleatoria ("estadístico").

- Se puede mostrar que, bajo ciertas suposiciones, χ^2_{obs} tiene una distribución χ^2 con $\nu = N - n$ grados de libertad.

La función de densidad de probabilidad χ^2 con diferentes ν



Aspectos estadísticos de la regresión lineal

La función densidad de probabilidad de la distribución χ^2 es

$$f_{\chi^2}(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{-1+\nu/2} e^{-x/2}.$$

- La **prueba** χ^2 nos califica estadísticamente el valor de los parámetros que se obtuvieron por medio de mínimos cuadrados.
- En esta prueba, calculamos χ^2_{obs} y la comparamos con la distribución teórica χ^2 con $\nu = N - n$ grados de libertad.

Aspectos estadísticos de la regresión lineal

La probabilidad de obtener un valor χ^2 grande o mayor que el observado es

$$p = \int_{\chi^2_{\text{obs}}}^{\infty} f_{\chi^2}(x)dx.$$

Que se conoce como la **prueba de valor-**p.

- Cuando los errores en los datos son independientes y distribuidos normalmente,
- y el modelo matemático “es correcto”,
- se puede mostrar que el valor-p será uniformemente distribuido entre 0 y 1.

En la práctica, valores-p particulares que son cercanos a los extremos de este intervalo nos indican que una o más suposiciones no son correctas.

Aspectos estadísticos de la regresión lineal

Aquí tenemos los tres casos generales.

1. Los valores-p están en $[0, 1]$. Nuestra solución produce un ajuste aceptable y nuestras hipótesis estadísticas sobre los errores de los datos son consistentes.

2. Los valores-p son muy pequeños.
   (a) Observaciones poco probables.
   (b) El modelo matemático $Ax = d$ es incorrecto.
   (c) La dispersión de los errores está pobremente estimada o bien los errores no están distribuidos de forma normal.

3. Los valores-p son muy grandes. El ajuste entre el modelo y los datos es casi exacto. Se pueden haber sobreestimado los errores en los datos. Una posibilidad es el cuchareo de los datos.

Aspectos estadísticos de la regresión lineal

- Por lo general, en problemas con muchos grados de libertad ν, el valor esperado de χ^2 se aproxima precisamente a este valor. (Teorema del límite central).

$$\begin{aligned} \langle f_{\chi^2} \rangle =& \nu, \\ \sigma =& \sqrt{2\nu}. \end{aligned}$$

- Además de analizar la distribución de χ^2_{obs}, es importante examinar los residuos correspondientes al modelo.

- Estos residuos deberían estar distribuidos de forma normal aproximadamente y no deberían tener algún patrón de distribución.

- Tales residuos pueden revelar el error en el modelo.

Aspectos estadísticos de la regresión lineal: covarianzas

- Estimadores de parámetros que se obtienen vía regresión lineal, son combinaciones de los datos.

- Si los errores en los datos están distribuidos normalmente, la estimación de parámetros también lo será, debido a que una combinación lineal de variables aleatorias normalmente distribuidas es normalmente distribuida.

- Para derivar el mapeo entre los datos y las covarianzas, consideremos un vector de datos d, de variables aleatorias normalmente distribuidas e independientes, operando sobre una transformación lineal general especificada por una matriz A.

- El mapeo de covarianza está definido por:

$$\mathrm{Cov}(Ad) = A\,\mathrm{Cov}(d)A^T.$$

Aspectos estadísticos de la regresión lineal: covarianzas

- La solución de mínimos cuadrados tiene $x_{L_2} = (A_w^T A_w)^{-1} A_w$. Y dado que los datos con peso tienen una matriz covariante identidad, la covarianza de los parámetros es

$$\mathrm{Cov}(x_{L_2}) = (A_w^T A_w)^{-1} A^T I_N A_w (A_w^T A_w)^{-1} = (A_w^T A_w)^{-1}$$

- En el caso de errores en los datos independientes e idénticamente distribuidos normalmente, la matriz de covarianza $\mathrm{Cov}(d)$ es simplemente la varianza σ^2 veces la matriz identidad de N por N; $i.e.$, I_N, por lo que:

$$\mathrm{Cov}(x_{L_2}) = \sigma^2 (A^T A)^{-1}.$$

- Note que en general, la matriz de covarianzas no es diagonal, indicando que **los parámetros están correlacionados**.
- Debido a que los resultados de los mínimos cuadrados son combinaciones lineales de los elementos del vector de datos, la dependencia estadística de los elementos en los parámetros x no es una sorpresa.

Aspectos estadísticos de la regresión lineal: covarianzas

- El valor esperado de la solución de mínimos cuadrados de $A_w x = d_w$ es

$$\langle x_{L_2} \rangle = (A_w^T A_w)^{-1} A_w^T \langle d_w \rangle.$$

- Y debido a que $\langle d_w \rangle = d_{\mathsf{v}}$ (verdadero), y que $A_w x_v = d_{vw}$, tenemos

$$A_w^T A_w x_v = A_w^T d_v.$$

- Así

$$\langle x_{L_2} \rangle = (A_w^T A_w)^{-1} A_w^T A_w x_v = x_v$$

- En estadśitica se dice que la solución de mínimos cuadrados no está **sesgada**.

Aspectos estadísticos de la regresión lineal: covarianzas

- Podemos calcular los intervalos de confianza del 95 % para cada parámetro, usando el hecho de que cada x_i tiene una distribución normal, con media dada por el elemento de m_v y varianza $\mathrm{Cov}(x_{L_2})_{i,i}$

- Los intervalos de confianza del 95 % están dados por

$$x_{L_2} \pm 1.96 \times \mathrm{diag}\left(\sqrt{\mathrm{Cov}(x_{L_2})}\right)$$

donde el número 1.96 aparece a partir de

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-1.96\sigma}^{+1.96\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz.$$

Aspectos estadísticos de la regresión lineal: Ejemplo 1

Regresión lineal sobre observaciones balísticas (datos sintéticos):

$$y(t) = m_1 + m_2 t - \frac{1}{2} m_3 t^2.$$

La variable dependiente $y(\cdot)$ representa mediciones de elevación, y t el tiempo, m_1 es la posición inicial, m_2 la velocidad inicial y $-m_3$ la aceleración de la gravedad (que apunta hacia la Tierra).

- Consideremos un conjunto $N = 10$ de datos sintéticos distribuidos normalmente e independientes, con una desviación estándar $\sigma = 8$ m.
- Generamos los datos con el siguiente vector de parámetros:

$$m_v = \begin{bmatrix} 10 & m \\ 100 & m/s \\ 9.8 & m/s^2 \end{bmatrix}$$

Aspectos estadísticos de la regresión lineal: Ejemplo 1

- Construimos a la matriz A, cuyo i-ésimo renglón es

$$A_{i,\cdot} = \left[1, t_i, -\frac{1}{2}t_i^2\right],$$

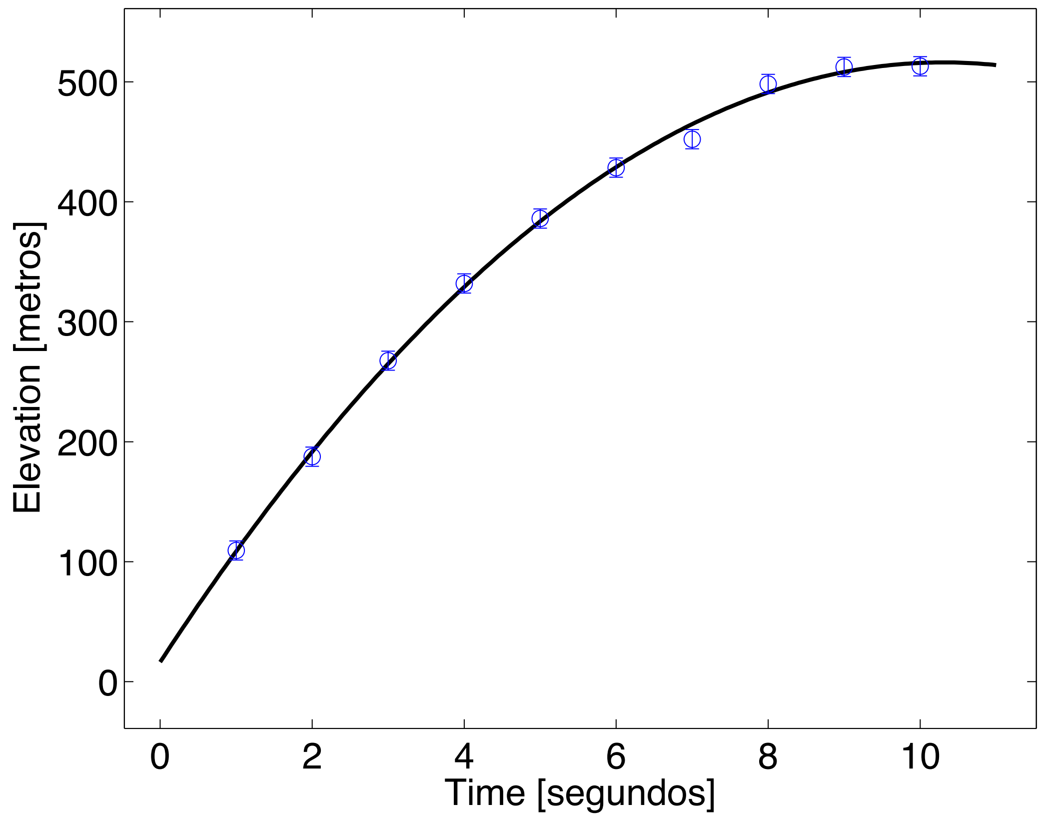
tal que

$$A = \begin{bmatrix} 1.0 & 1.0 & -0.5 \\ 1.0 & 2.0 & -2.0 \\ 1.0 & 3.0 & -4.5 \\ 1.0 & 4.0 & -8.0 \\ 1.0 & 5.0 & -12.5 \\ 1.0 & 6.0 & -18.0 \\ 1.0 & 7.0 & -24.5 \\ 1.0 & 8.0 & -32.0 \\ 1.0 & 9.0 & -40.5 \\ 1.0 & 10.0 & -50.0 \end{bmatrix}$$

Aspectos estadísticos de la regresión lineal: Ejemplo 1

Obtenemos los parámetros por mínimos cuadrados empleando las ecuaciones normales con pesos:

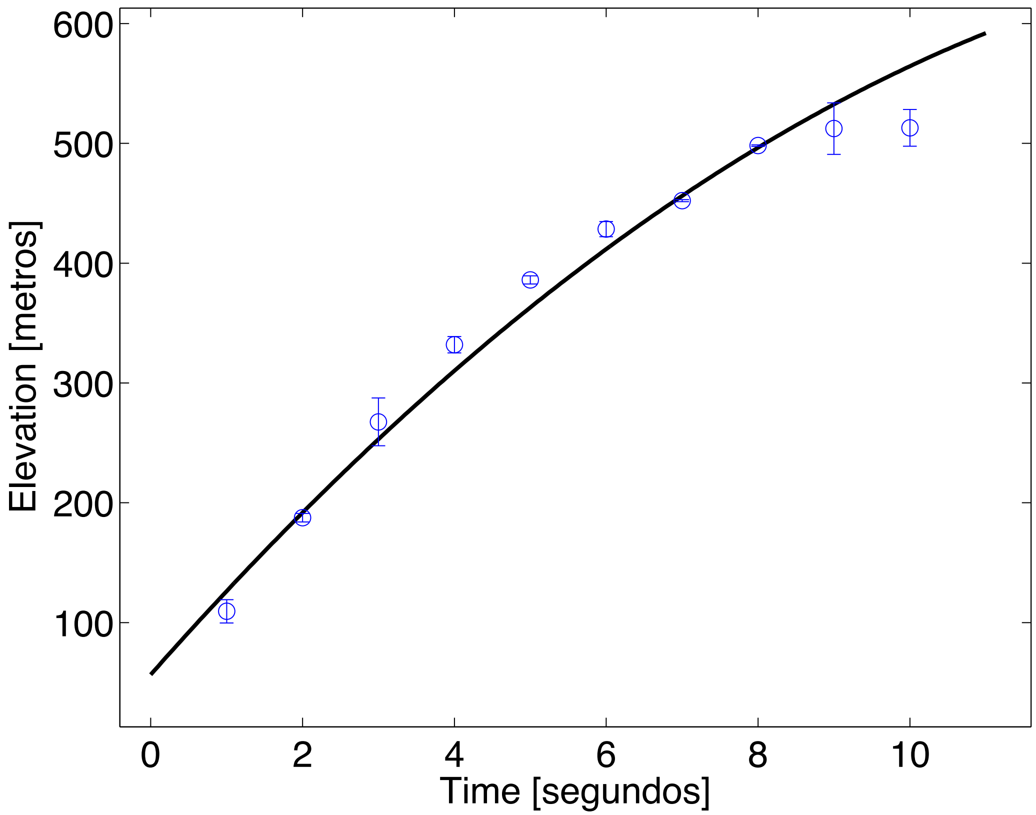
$$m_{L_2} = \begin{bmatrix} 16.4 & m \\ 97.0 & m/s \\ 9.4 & m/s^2 \end{bmatrix}$$



Aspectos estadísticos de la regresión lineal: Ejemplo 1

Mismos datos, pero con una dispersión aleatoria σ `= 10.0*randn(length(t),1);`

$$m_{L_2} = \begin{bmatrix} 56.5 & m \\ 71.8 & m/s \\ 4.2 & m/s^2 \end{bmatrix}$$



Aspectos estadísticos de la regresión lineal: Ejemplo 1

- Los valores de la **matriz de covarianza**, del primer caso, es:

$$\mathrm{Cov}(m_{L_2}) = \begin{bmatrix} 88.5333 & -33.6000 & -5.3333 \\ -33.6000 & 15.4424 & 2.6667 \\ -5.3333 & 2.6667 & 0.4848 \end{bmatrix}$$

- Los intervalos de confianza (95 %) son:

$$m_{L_2} = \begin{bmatrix} 16.4 \pm 18.4 \\ 97.0 \pm 7.7 \\ 9.4 \pm 1.4 \end{bmatrix}$$

- El valor de χ^2 de esta regresión es 4.2048, y el número de grados de libertad $\nu = N - n = 10 - 3 = 7$, tal que el valor-p es

$$p = \int_{4.2}^{\infty} \frac{x^{5/2} e^{-x/2}}{2^{7/2}\Gamma(7/2)} dx \approx 0.76$$

Aspectos estadísticos de la regresión lineal

- Si consideramos combinaciones entre los valores paramétricos, la interpretación de las incertidumbres se vuelven más complicada.

- Para caracterizar la incertidumbre que proviene de modelo de forma efectiva, examinaremos los **intervalos de confianza** por pares de parámetros.

- Cuando las **regiones de confianza conjuntas** de los parámetros se proyectan sobre los ejes coordenados m_i, se pueden obtener intervalos de confianza significativamente mayores que si se consideraran de forma individual.

Aspectos estadísticos de la regresión lineal: distribución normal

$$f(m) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(C)}} \exp\left(-\frac{(m-\mu)^T C^{-1}(m-\mu)}{2}\right)$$

- Supongamos que el **vector de parámetros estimados**, $m \in \mathbb{R}^n$, está distribuido según una **normal multivariada** de dimensión n, con matriz de covarianza C, entonces se puede mostrar que la variable aleatoria

$$(m - m_{L_2})^T C^{-1} (m - m_{L_2})$$

posee una distribución χ^2 con n grados de libertad.

Aspectos estadísticos de la regresión lineal: distribución normal

En estadística, un **percentil** es el valor de una variable abajo del cual se encuentran un cierto porcentaje de las mediciones.

- Si denotamos por Δ^2 el percentil 95 % de la distribución χ^2 con n grados de libertad, entonces se define la **región de confianza del 95 %** mediante la desigualdad

$$(m - m_{L_2})^T C^{-1} (m - m_{L_2}) \leq \Delta^2$$

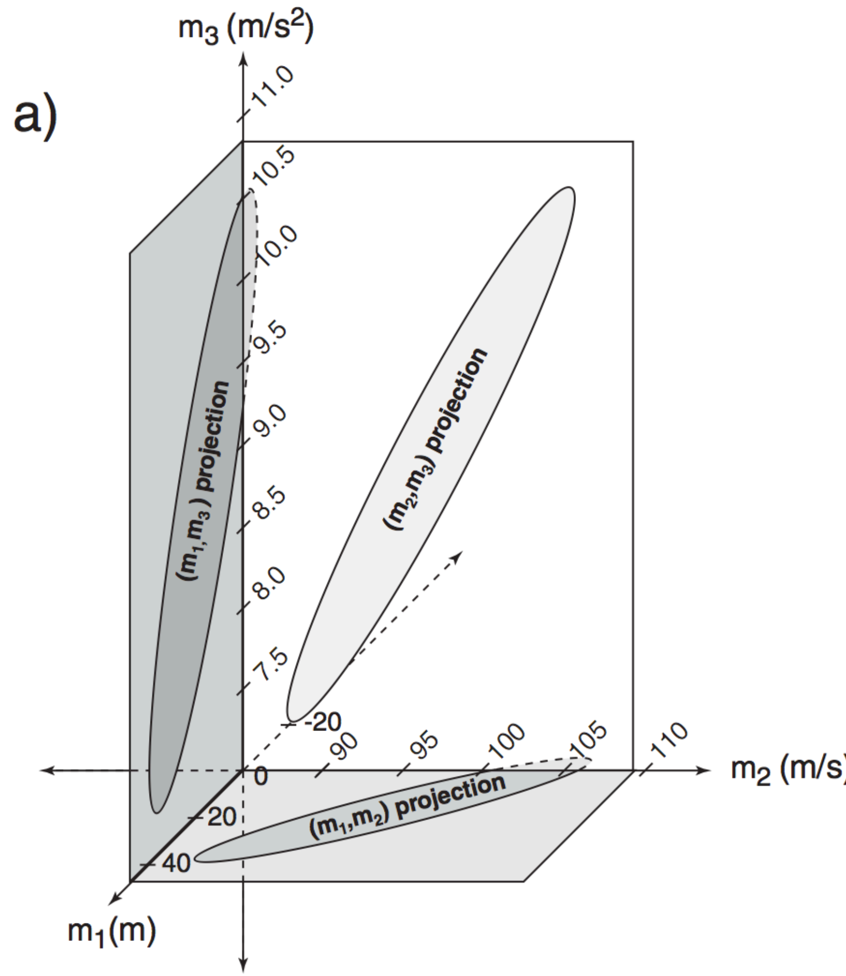
- La región de confianza, definida por esta desigualdad, resulta ser un **elipsoide** n-**dimensional**. En donde sus **ejes** se corresponden con las direcciones de los **valores propios** de C y sus **valores propios** a los cuadrados de los **semiejes**.

Aspectos estadísticos de la regresión lineal: distribución normal

- Del elipsoide n-dim deseamos hallar sus proyecciones sobre espacios de dimensión inferior.

- Para ello tomamos sólo aquellos renglones y columnas de C y los respectivos elementos de m, que corresponden a la dimensión que queremos explorar.

- En este caso, el número de grados de libertad en el cálculo asociado de χ^2, se debería reducir a el número de parámetros en la proyección.



Casa abierta al tiempo

Aspectos estadísticos de la regresión lineal: distribución normal

- Dado que la matriz de covarianza y su inversa son simétricas y positivas definidas, podemos diagonalizar a C^{-1}:
$$C^{-1} = P^T \Lambda P,$$
en donde Λ es una matriz diagonal de los valores propios y las columnas de P son vectores propios ortonormales.

- Los semiejes definidos por las columnas de P se definene como **ejes principales del elipsoide de error**, donde la dirección del i-ésimo semieje mayor está dado por $P_{\cdot,i}$, y tiene una longitud igual a
$$\frac{\Delta}{\sqrt{\Lambda_{i,i}}}.$$

Aspectos estadísticos de la regresión lineal: distribución normal

- Debido a que la matriz de covarianza de los parámetros no es diagonal, los ejes principales no están perfectamente alineados con las direcciones de los ejes m_i.
- Sin embargo, podemos proyectar al elipsoide de confianza apropiado sobre los ejes m_i y obtener una "caja" que incluya al 95 % del elipsoide de error, con un volumen adicional externo.
- Tal caja nos provee de un intervalo de confianza conservador para un familia de valores paramétricos conjunta.

Aspectos estadísticos de la regresión lineal: distribución normal

- Las correlaciones entre pares de parámetros (m_i, m_j) son una medida de la inclinación de los elipsoides de error respecto a los ejes de los parámetros.

1. Una correlación próxima a $+1$ significa una alta excentricidad (pendiente positiva) respecto al eje principal mayor.

2. Una correlación cero significa que la proyección tiene ejes principales alineados con los ejes del plano (m_i, m_j).

3. Y una correlación cercana a -1 significa que la proyección es altamente excéntrica con pendiente negativa a lo largo del eje principal.

Elipsoides de error: Ejemplo 2

- Correlaciones paramétricas

$$\rho_{m_i,m_j} = \frac{\mathrm{Cov}(m_i, m_j)}{\sqrt{\mathrm{Var}(m_i)\,\mathrm{Var}(m_j)}}$$

$$\rho_{m_1,m_2} = -0.91$$

$$\rho_{m_1,m_3} = -0.81$$

$$\rho_{m_2,m_3} = 0.97$$

Los tres parámetros son altamente dependientes estadísitcamente y los elipsoides de error son excéntricos e inclinados.

Elipsoides de error: Ejemplo 2

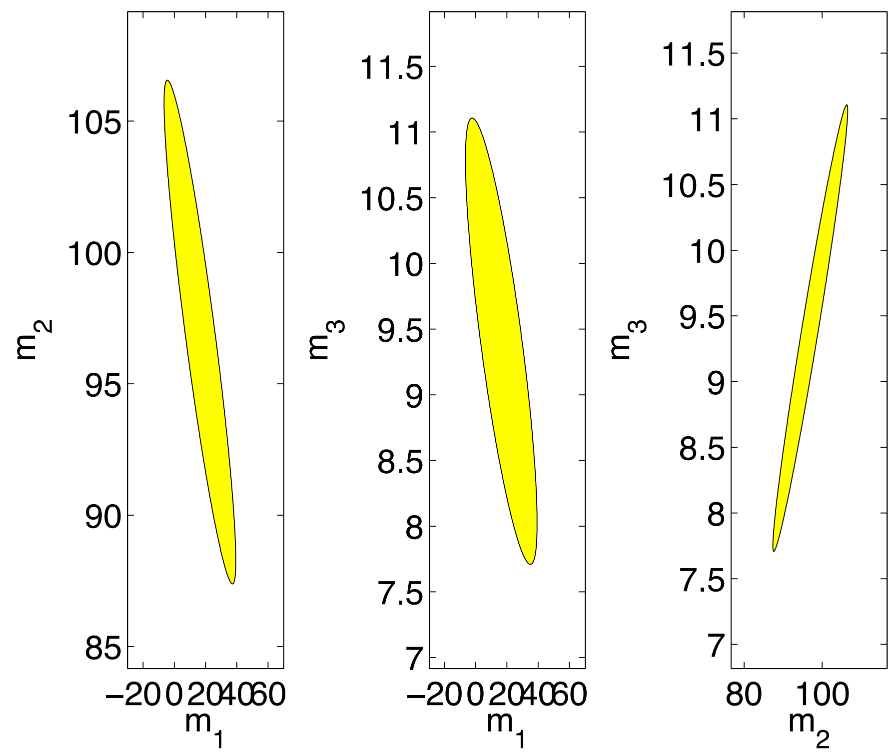


Figura 1: Proyecciones del 95 % del elipsoide de error sobre los ejes de parámetros. Proyecciones sobre planos.

Elipsoides de error: Ejemplo 2

- Diagonalización de de C^{-1}

$$P = \begin{bmatrix} \vdots & \vdots & \vdots \\ P_{\cdot,1} & P_{\cdot,2} & P_{\cdot,3} \\ \vdots & \vdots & \vdots \end{bmatrix} \approx \begin{bmatrix} -0.03 & -0.93 & 0.36 \\ -0.23 & 0.36 & 0.90 \\ 0.97 & 0.06 & 0.23 \end{bmatrix}.$$

- Valores propios

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \approx \begin{bmatrix} 104.7 \\ 0.0098 \\ 0.4046 \end{bmatrix}.$$

Elipsoides de error: Ejemplo 2

- Longitudes de los semiejes elipsoidales de confianza de 95 %:

$$\sqrt{F^{-1}_{\chi^2,3}}\begin{bmatrix}1/\sqrt{\lambda_1}\\1/\sqrt{\lambda_2}\\1/\sqrt{\lambda_3}\end{bmatrix}\approx\begin{bmatrix}0.24\\24.72\\3.85\end{bmatrix}.$$

- Donde $F^{-1}_{\chi^2,3}(0.95) \approx 2.80$ es el 95-avo percentil de la distribución χ^2 con 3 grados de libertad.

Elipsoides de error: Ejemplo 2

- Intervalos de confianza del 95 % (proyecciones sobre los ejes de m_1, m_2, m_3):

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} 16.42 \pm 23.03\,\mathsf{m} \\ 96.97 \pm 9.62\,\mathsf{m/s} \\ 9.41 \pm 1.70\,\mathsf{m/s}^2 \end{bmatrix}$$

Desviaciones estándar desconocidas

- Anteriormente supusimos que conocíamos la desviación estándar σ de las mediciones.
- Supongamos que no poseemos esta información de σ a priori.
- Supongamos, también, que los errores de medición son **independientes** y que están **distribuidos normalmente**, con valor esperado igual a 0 y que poseen una presunta σ.
- Entonces, obtenemos la regresión lineal y **estimamos** a σ a partir de los **residuos**.

Desviaciones estándar desconocidas

- Primero hallamos la solución de mínimos cuadrados al problema sin pesos $Ax = d$, con residuos $r = d - Ax_{L_2}$.

- Estimamos la desviación estándar a partir de los residuos

$$s = \sqrt{\frac{1}{N-n}\sum_{i=1}^{N} r_i^2}.$$

- Como es de esperarse, tendremos que pagar el costo de no conocer la "verdadera" desviación estándar.

Desviaciones estándar desconocidas

- Si las desviaciones estándar fueran conocidas a priori, entonces los errores en los parámetros serían
$$x_i' = \frac{x_i - x_{v_i}}{\sigma}$$
teniendo una **distribución normal estándar**.
- Si en vez de conocer una σ la estimamos, s, entonces los errores en los parámetros son
$$x_i' = \frac{x_i - x_{v_i}}{s}$$
teniendo una **distribución** t**-Student** con $\nu = N - n$ grados de libertad.
- Para muy pequeños grados de libertad ν, lo anterior produce intervalos de confianza mucho más anchos.
- Pero cuando ν es muy grande, s se convierte cada vez en un mejor estimador de σ.

Desviaciones estándar desconocidas

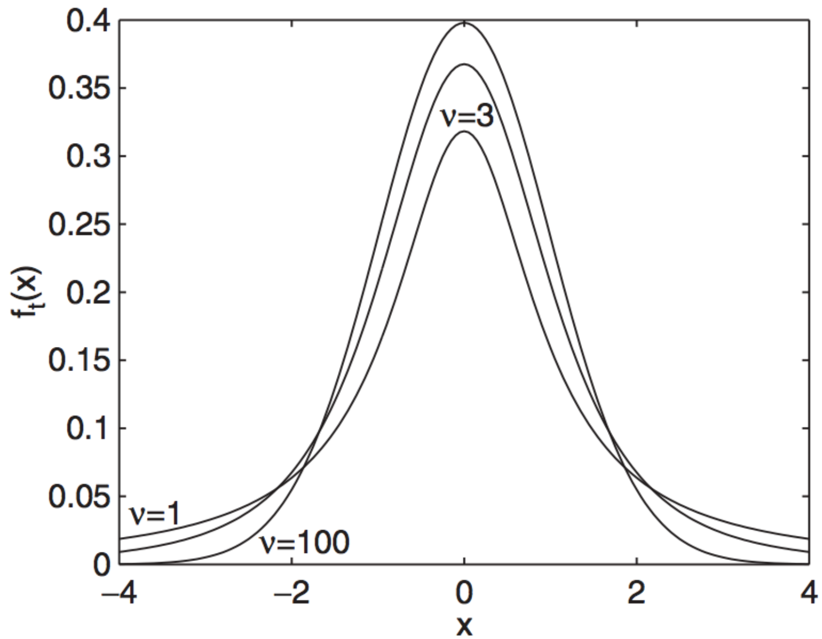


Figura 2: La función de densidad de probabilidad t de Student para $\nu = 1, 3, 100$.

$$f_t(x) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{1}{\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

Desviaciones estándar desconocidas: Ejemplo 2.3

Regresión L_1

Casa abierta al tiempo