Estimació d'una variable aleatòria a partir d'una altra

Plantejament del problema:

Suposem que tenim dues v.a. X i Y distribuïdes de manera conjunta i que, per a un resultat concret de l'experiment aleatori, coneixem el valor de la v.a. X: $X = x$. A partir d'aquesta dada volem 'endevinar' el valor de Y.

Per a cada valor $X = x$, Y pot prendre tot un conjunt de valors (ho denotam $\Omega_{Y|X=x}$, veure la figura 1-esquerra). Quin és el 'millor' que podem triar? Per poder decidir si un valor és 'millor' o 'pitjor' que un altre haurem de definir un criteri. Per exemple, podem considerar que el 'millor' valor és aquell que fa que l'error que cometrem si ens equivocam sigui el menor possible.

Si hem triat un valor y^* i ha sortit un valor y, l'error de l'estimació ho podem calcular com: $(y^* - y)^2$. Per a cada resultat y tendrem un valor possible de l'error. L'error mitjà que cometrem es calcularà com:

$$\mathrm{ECM} = E((Y^* - Y)^2)$$

on Y^* denota la variable aleatòria que ens dóna el valor de Y estimat a partir de X, la v.a. Y representa tots els valors possibles de y, i per tant l'ECM representa l'**error quadràtic mitjà** de l'estimació. El que volem és trobar una v.a. Y^* que minimitzi l'anterior expressió.

Es pot demostrar que el problema té les següents solucions:

- **Millor estimació de Y a partir de X**: $Y^* = E(Y|X)$

  (aquesta funció rep el nom de **corba de regressió** de la mitjana de Y sobre X).

- **Millor estimació <u>lineal</u> de Y a partir de X**: $Y^* = aX + B$, on $a = \frac{\sigma_{XY}}{\sigma_X^2}$ i $b = \mu_Y - a\mu_X$

  (aquesta recta rep el nom de **recta de regressió** lineal de Y sobre X).

La següent figura mostra un exemple de cada tipus d'estimació.

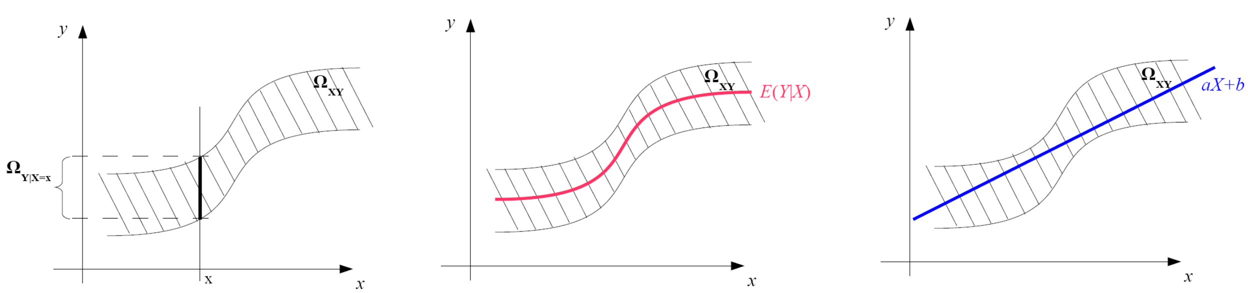

Figura 1: Esquerra: exemple de conjunt de valors possibles de Y donat un valor $X = x$. Centre: millor estimació de Y a partir dels valors de X. Dreta: millor estimació lineal de Y a partir dels valors de X. Suposam distribució uniforme de la probabilitat.

Per mesurar la qualitat de l'estimació utilitzam la raó de correlació. La **raó de correlació de Y sobre X** es defineix com:

$$\eta^2_{Y|X} = \frac{\mathrm{Var}(Y^*)}{\mathrm{Var}(Y)}$$

Propietats:

- $0 \leq \eta^2_{Y|X} \leq 1$

  El valor 1 significa que l'estimació és perfecta ($Y^* = Y$ per a tot x).

- si $Y^* = E(Y|X)$ llavors $\eta^2_{Y|X} = 1 - \frac{\mathrm{ECM}}{\mathrm{Var}(Y)}$

Exemple 24:

Estimau la trajectòria lineal d'un avió de guerra a partir de les posicions de caiguda de les bombes que ha llançat. Les coordenades dels impactes es mostren en la taula següent:

x	y
0.5	10
0.4	12
0.9	14
1.5	11
1.8	14.5
2	17
2.5	13

x	y
3	16
3.2	17
3.4	14
3.7	16
4	19
4.5	17
5.5	21

Exemple 25:

(Exercici 24). $E(Y|X)$ és la funció $g(X)$ que "millor"aproxima Y on "millor" indica que l'error quadràtic mitjà $E\left((Y-g(X))^2\right)$ és mínim. Aquesta funció de X s'anomena la **corba de regressió de la mitjana de Y sobre X**. *Anàlogament tenim la corba de regressió de la mitjana de X sobre Y. Determinau les corbes de regressió de les mitjanes i les rectes de regressió lineal si (X, Y) està distribuït uniformement en el triangle de vèrtexs (0,0), (1,0) i (1,2).*

Problemes proposats: 17ef, 25a, 16d