Frontera de Eficiencia de Futuros de Energía Eléctrica

Seminario DERIVEX

Octubre 24 de 2012

Quantil | Matemáticas Aplicadas

Plan de la presentación

La Aproximación Bayesiana a la Estadística

- La teoría clásica utiliza la información muestral para hacer inferencias sobre los parámetros de interés.
- Usualmente existe información incial sobre los parámetros de un modelo. Los parámetros del modelo pueden ser retornos esperados.
- Permite condicionar a los datos observados. En el análisis clásico se promedia sobre los los datos, aun los no observados.
- Distribuciones exactas. La teoría clásica se basa en muchas ocasiones en teoría asintótica.
- Coherencia y racionalidad: La teoría Bayesiana es una aproximación general al problema de inferencia consistente con la teoría de la decisión.

La Aproximación Bayesiana a la Estadística

- Las reglas de decisión en un contexto Bayesiano son óptimas desde un punto de vista clásico.
- Mécanica Bayesiana: Siempre se sabe qué hacer.
- Computacionalmente es difícil.
- Razones técnicas:
  1. Permite hacer inferenecia estadístca en modelos no regulares.
  2. Permite introducir incertidumbre en los parámetros para hacer prediciciones.
  3. Permite hacer pruebas de modelos no anidados.
  4. Se pueden analizar modelos jerárquicos de forma conceptualmente muy coherente.

- Probabilidad = Incertidumbre. En la teoría Bayesiana el concepto de probabilidad tiene una interpretación distinta a la teoría clásica o frecuentista. El concepto de probabilidad es una medida de la incertidumbre sobre la ocurrencia de un evento.
- A diferencia de la teoría clásica es posible dar interpetaciones sobre la incertidumbre de un parámetro que no están basadas en la repetición bajo condiciones iguales de un experimento (intervalos de confianza).
- Por ejemplo, es posible cuantificar en términos probabilísticos el grado de incertidubre con la cuál se hace un pronóstico.

El concepto de probabilidad

- Existen por lo menos tres interpretaciones del concepto: objetiva (Fermat, Pascal, Huygens, etc), subjetiva (Ramsey, de Finetti, Savage), lógica.
- Axiomas de Kolmogorov.

El concepto de probabilidad

- Riesgo e incertidumbre: La paradoja de Ellsberg: Una urna contiene 90 bolas donde 30 son rojas. El resto de las bolas son amarillas o negras y su distribución es desconocida.
  1. Apuesta A: Quien saque una bola roja gana el resto pierden.
  2. Apuesta B: Quien saque una bola amarilla gana, el resto pierde.
- La mayoría de las personas optan por la A.
- Ahora consideremos la apuesta:
  1. Apuesta C: Quien saque una bola roja o negra gana
  2. Apuesta D: Quien saque una bola amarilla o negra gana,
- La mayoría de las personas escogen D.

- Ellsberg explica éste resultado en términos de la diferencia entre el riesgo e incertidumbre. Las personas sometidas a estas escogencias suponen prudentemente que la distribución desconocida entre bolas rojas y amarillas pueden traerles desventaja y por lo tanto escogen en ambas ocasiones bajo el riesgo conocido (1/3 en la primera prueba, 2/3 en la segunda).
- Llama la atención sobre la necesidad de una teoría para modelar la incertidumbre.

El concepto de probabilidad

- Una forma de interpretar el concepto de probabilidad desde un punto de vista lógico es de acuerdo al concepto de razonamiento plausible (Jaymes): cuano en el mundo real observamos un evento B que típicamente es consecuencia de un evento A, decimos que A es plausible pero usalmente no es posible deducir que A ha sucedido. La idea predominante es que A es plausible dado que observamos B. La lógica es que si B ha ocurrido, esto arroja evidencia en favor de A.
- Al fundamentar la teoría de la probabilidad de esta forma se obtiene una formalización de la idea del grado de incertidumbre sobre la ocurrencia de un evento (plausibilidad del evento).

El concepto de probabilidad

- Luego la interpretación de la probabilidad de un evento no está relacionada con la frecuencia de un evento repetido sino con el grado de incertidumbre del evento. Esta es la interpreatción subjetivista del concepto de probabilidad.
- Para de Finetti la probabiidad (objetiva) de un evento no existe. Es algo tan ficticio y en contradición con laevidencia cuanto la existencia del éter.
- Cuál es la probabilidad de cada número cuando se lanza un dado al aire?
- La idea de aprendizaje en un ambiente incierto puede ser sutil.

- La paradoja del gato I. Una persona está frente a tres puertas cerradas. Se sabe que detrás de alguna de las puertas hay un gato. La persona se le pide escoger una puerta. Antes de abrir cualquier puerta, una segunda persona que sabe exactamente que hay detrás de cada puerta y conoce también cuál fue la puerta elegida por la primera persona, éste abre una de las puertas que sea la elegida por la primera persona y en la que no esté el gato. Ahora, con una puerta abierta en la que no está el gato, se le pregunta a la primera persona si desearía cambiar de puerta.

El concepto de probabilidad

- El sentido común dice que no hace diferencia. Pero la teoría de la probabilidad dice otra cosa. La probabilidad de encontrar el gato en alguna de las dos puertas al cambiar la elección original es mayor que la probabilidad de que el gato esté en la primera puerta elegida.
- Llama la atención sobre la necesidad de una teoría de aprendizage.

El concepto de probabilidad

- Definición probabiidad condicional. Dados dos evento A y B, tal que $P(B) > 0$ definimos la probabilidad condicional de A dado B como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1}$$

- El teorema de Bayes (o regla de Bayes) afirma que:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}. \tag{2}$$

- Este resultado es la base de toda la estadística Bayesiana.

- La paradoja del gato II: Para formalizar este problema, supongamos que la primera elección fue la tercera puerta. Sean A_1, A_2 y A_3 los eventos en los cuales el gato está detrás de la puerta $1, 2$ o 3 respectivamente. Sean B_1 y B_2 los eventos en los cuales el segundo jugador abre la puerta 1 o 2 reespectivamente. Nuetro objetivo es calcular $P\left(A_i \left| B_j\right.\right)$. Entonces dada la información del problema es natural suponer:

$$P(A_i) = \frac{1}{3}, P\left(B_1 \left| A_1\right.\right) = P\left(A_2 \left| B_2\right.\right) = 0$$

$$P(B_1 \left| A_2\right.) = P(B_2 \left| A_1\right.) = 1$$

y

$$P(B_1 \left| A_3\right.) = P(B_2 \left| A_3\right.) = \frac{1}{2}.$$

Entonces si la segunda persona abre la puerta 2 es fácil calcular, usando la regla de Bayes, $P\left(A_1 \left| B_2\right.\right) = \frac{2}{3}$.

Implementación

- Sea Θ un espacio de parámetros o estados de la naturaleza y Ξ un espacio de datos observables.
- En términos de funciones de densidad el teorema se puede expresar como:

$$\mathrm{f}(\theta\,|y) = \frac{f(y|\theta)f(\theta)}{f(y)}$$

donde $f(y)$ es la distribución marginal de la variable aleatoria Y (o distribución marginal de los datos):

$$f(y) = \int_{\Theta} f(y\,|\theta)f(\theta)d\theta,$$

Implementación

- $f(\theta\,|y)$ es la distribución expost (posterior) del parámetro θ dado los datos observados y, $f(y\,|\theta)$ es la distribución muestral.

Implementación

- La función $\mathcal{L}(\theta|y) = f(y|\theta)$, como función de θ se llama la función de verosimilitud.
- $f(\theta)$ es la distribución inicial (prior) sobre los parámetros.
- Obsérvese que no se ha hecho ninguna hipótesis sobre la forma de la distribución muestral. En general suponemos que Y es un vector de obsrevaciones y $f(y|\theta)$ es la distribución conjunta o distribución del vector aleatorio Y.

El modelo de Black y Litterman con Reducción de Dimensión

- Suponga que el espacio de parámetros son los factores de riesgo de mercado X. Por ejemplo los retornos logarítmicos de un conjunto de acciones (diferentes puntos de la curva forward del precio de electricidad).
- En general vamos a modelar X mediante un modelo de factores: $X_t = B_t F_t + \epsilon_t$
- Es importante que el modelo *estático* anterior se pueda estimar razonablemente, que se pueda hacer un buen modelo *dinámico* de F_t y que B_t sea una matriz de carga determinística.

El modelo de Black y Litterman con Reducción de Dimensión

- Sea $f(x)$ la distribución inicial de X. La idea es deducir esta distribución del modelo de factores.
- El administrador tiene opiniones que expresa como una variable aleatoria V condicional a las observaciones del parámetro X.

El modelo de Black y Litterman con Reducción de Dimensión

- Supongamos que $X \sim N(\mu, \Sigma)$, $g(X) = PX$ donde P es una matriz que selecciona los activos (o combinación) en los cuales el experto tiene una expectativa.
- $V \mid g(X) \sim N(g(X), \Omega)$, donde Ω representa la confianza que el experto tiene en sus expectativas. La distribución $V \mid g(X)$ es la distribución muestral. La interpretación es que dado $g(X)$ el experto genera sus expectativas usando una distribucion $N(g(X), \Omega)$. Un caso particular es cuando $\Omega = \left(\frac{1}{c} - 1\right) P\Sigma P^T$, c un escalar positivo.
- Obsérvese que $P\Sigma P^T$ significa que el experto evalua la incertidumbre de $g(X)$ utilizando la distribución estimada de los factores de riesgo.

El modelo de Black y Litterman con Reducción de Dimensión

- El factor $\left(\frac{1}{c} - 1\right)$ lo que hace es disminuirla o aumentarla proporcionalmente.
- Si $c \to 0$ el experto tiene expectativas muy inciertas (la confianza en el experto es baja y la prior y posterior son similares). Entre más cerca a 1 esté c, menos inciertas son las expectativas del experto (la confianza en el experto es mayor y la prior y posterior son menos parecidas).
- En conclusion, cuando la confianza en el experto es alta, la prior y la posterior se distancian y cuando es baja la prior y la posterior son parecidas.

El modelo de Black y Litterman con Reducción de Dimensión

- Supongamos que el experto tiene expectativas $V = v$. Obsérvese que las expectativas son sobre un combinación lineal de los factores de riesgo.
- Se puede demostrar que:

$$X \mid v \sim N(\mu_{BL}, \Sigma_{BL})$$

donde:

$$\mu_{BL}(v, \Omega) = \mu + \Sigma P^T (P\Sigma P^T + \Omega)^{-1}(v - P\mu) \quad (3)$$

$$\Sigma_{BL} = \Sigma - \Sigma P^T (P\Sigma P^T + \Omega)^{-1} P\Sigma. \quad (4)$$

Obsérvese que en este caso particular Σ_{BL} no depende de v.

El modelo de Black y Litterman con Reducción de Dimensión

- Ahora si $F \sim N(\mu_F, \Sigma_F)$ entonces $X \sim N(B\mu_F, B^T\Sigma_F B)$ desconociendo el papel de ϵ.
- Si $V = PX$ entonces $V = PBF$ desconociendo el papel de ϵ. Supongamos que los views se expresar de la forma: $V \mid PBF \sim N(PB\mu_F, (\frac{1}{c} - 1)(PB)^T\Sigma_F PB)$.
- Entonces $V \mid PBF \sim N(\widetilde{P}\mu_F, \widetilde{\Omega})$ donde $\widetilde{\Omega} = (\frac{1}{c} - 1)(PB)^T\Sigma_F PB$.
- Luego el modelo BL en términos de la distribución de los factores y con views sobre las variables observables X es:

El modelo de Black y Litterman con Reducción de Dimensión

Sustituir P por $\widetilde{P}$, Ω por $\widetilde{\Omega}$ y donde v es un view sobre PX

$$F \mid v \sim N(\mu_{BL}, \Sigma_{BL})$$

donde:

$$\mu_{BL}(v, \Omega) = \mu + \Sigma P^T (P \Sigma P^T + \Omega)^{-1} (v - P\mu) \tag{5}$$

$$\Sigma_{BL} = \Sigma - \Sigma P^T (P \Sigma P^T + \Omega)^{-1} P \Sigma. \tag{6}$$

El modelo de Black y Litterman con Reducción de Dimensión

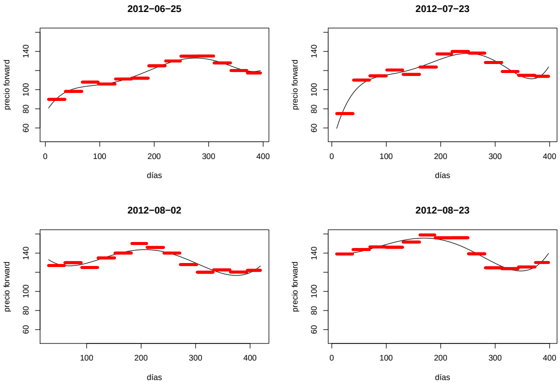
- Ahora con la distribución posterior para los factores $Fv \sim N(\mu_{BL}, \Sigma_{BL})$ se puede ahora calcular la distribución posterior para los factores de riesgo:

$$X \mid v \sim N(B\mu_{BL}, B^T \Sigma_{BL} B) \tag{7}$$

- Con esta distribución expost de los factores de riesgo construimos la frontera de eficiencia de los factores X.

Un modelo de factores semiparamétrico (Borak - Weron 2008)

- Curva forward típica que queremos modelar.



Un modelo de factores semiparamétrico (Borak - Weron 2008)

- En el modelo que se propone a continuación se mantiene la misma notación de Borak y Weron. Sin embargo, la forma de reintepretarlo para implementarlo como una aplicación del modelo de Black y Litterman con reducción de dimensión es:
    1. $Y \rightarrow X$
    2. $Z \rightarrow F$
    3. $X \rightarrow Z$
- Obsérvese que Z no aprece en el modelo de BL luego su introducción aquí es simplemente como covariantes adicionales (en este caso específico son variables determinísticas predecibles).

Un modelo de factores semiparamétrico (Borak - Weron 2008)

- El modelo propuesto es:

$$Y_{t,j} = m_0(X_j) + \sum_{l=1}^{L} Z_{t,l} m_l(X_j) + \epsilon_{t,j} \tag{8}$$

donde

- $Y_t = (Y_{t,1}, ..., Y_{t,J_t})$ es la observación en t de la curva forward con vencimientos $j = 1, ...J_t$.
- $Z_{t,l}$ son los factores comúnes (no observables).
- X_j son covariantes observables.
- $m_l(X_j)$ es un B-spline.
- Obsérvese que tiene la forma: $Y_t = B_t Z_t + \epsilon_t$, donde B_t son covariantes determinísticos en esta aplicación y Z_t son factores no observados que se estiman a continuación.

Un modelo de factores semiparamétrico (Borak - Weron 2008)

- Las funciones B-splines son de la forma:

$$m_l(X_j) = \sum_{k=1}^{K} a_{l,k}\varphi_k(X_j) \qquad (9)$$

donde K es el númerto de nodos, φ_k son funciones spline y $a_{l,k}$ son coeficientes.

Un modelo de factores semiparamétrico (Borak - Weron 2008)

Para determinar los factores no observables $Z_{t,l}$ y los coeficientes $a_{l,k}$ se resuleve el siguiente problema:

$$\sum_{t=1}^{T}\sum_{j=1}^{J}\left\{Y_{t,j}-\sum_{l=1}^{L}Z_{t,l}\sum_{k=1}^{K}a_{l,k}\varphi_k(X_j)\right\} \tag{10}$$

CP usando muchos predictores (Stock y Watson (2002))

- Se mantien la misma notación de los artículos originales. Para implementarlo como un modelo BL con reducción de dimensión cambiamos:
  1. $Y \to X$
  2. $X \to Z$

- Supóngase que Y_t es la serie a predecir y sea X_t un objeto de N series de tiempo las cuales se cree predicen Y_t. Ahora suponemos que el proceso de (X_t, Y_t) se puede representar como:

$$X_t = \Lambda F_t + e_t \tag{11}$$

$$Y_{t+h} = \beta_F F_t + \beta_w w_t + \epsilon_{t+h} \tag{12}$$

donde Λ es una matriz, β_F y β_w son vectores, e_t es un vector de $N \times 1$ errores indiosincráticos, w es un vector de m rezagos de Y_t, h es el horizonte del pronóstico y ϵ_{t+h} es el error del pronóstico.