Introduction to Time Series Analysis. Lecture 8.

Peter Bartlett

Forecasting

1. Linear prediction.
2. Projection in Hilbert space.
3. Forecasting and backcasting.
4. Prediction operator.

Review: least squares linear prediction

Consider a **linear predictor** of X_{n+h} given $X_n = x_n$:

$$f(x_n) = \alpha_0 + \alpha_1 x_n.$$

For a stationary time series $\{X_t\}$, the best linear predictor is $f^*(x_n) = (1 - \rho(h))\mu + \rho(h)x_n$:

$$\begin{aligned} \mathrm{E}\left(X_{n+h} - (\alpha_0 + \alpha_1 X_n)\right)^2 &\geq \mathrm{E}\left(X_{n+h} - f^*(X_n)\right)^2 \\ &= \sigma^2(1 - \rho(h)^2). \end{aligned}$$

Linear prediction

Given $X_1, X_2, \ldots, X_n$, the best linear predictor

$$X_{n+m}^n = \alpha_0 + \sum_{i=1}^n \alpha_i X_i$$

of X_{n+m} satisfies the **prediction equations**

$$\mathrm{E}\left(X_{n+m} - X_{n+m}^n\right) = 0$$

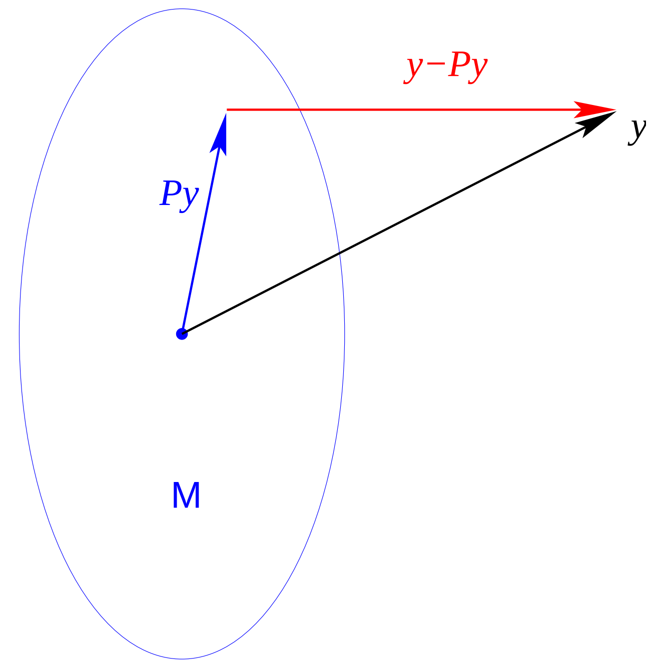
$$\mathrm{E}\left[\left(X_{n+m} - X_{n+m}^n\right) X_i\right] = 0 \qquad \text{for } i = 1, \ldots, n.$$

This is a special case of the *projection theorem*.

Projection Theorem

If $\mathcal{H}$ is a Hilbert space,
$\mathcal{M}$ is a closed linear subspace of $\mathcal{H}$,
and $y \in \mathcal{H}$,
then there is a point $Py \in \mathcal{M}$
(the **projection of** y **on** $\mathcal{M}$)
satisfying

1. $\|Py - y\| \leq \|w - y\|$ for $w \in \mathcal{M}$,
2. $\|Py - y\| < \|w - y\|$ for $w \in \mathcal{M}$, $w \neq y$
3. $\langle y - Py, w \rangle = 0$ for $w \in \mathcal{M}$.



Hilbert spaces

Hilbert space $=_{\text{complete}}$ inner product space:

Inner product space: vector space, with inner product $\langle a, b\rangle$:

- $\langle a, b\rangle = \langle b, a\rangle$,
- $\langle \alpha_1 a_1 + \alpha_2 a_2, b\rangle = \alpha_1\langle a_1, b\rangle + \alpha_2\langle a_2, b\rangle$,
- $\langle a, a\rangle = 0 \Leftrightarrow a = 0$.

Norm: $||a||^2 = \langle a, a\rangle$.

complete = limits of Cauchy sequences are in the space

Examples:

1. $\mathbb{R}^n$, with Euclidean inner product, $\langle x, y\rangle = \sum_i x_i y_i$.
2. $\mathcal{H} = \{$mean 0 random variables X: $\mathrm{E}X^2 < \infty\}$,
with inner product $\langle X, Y\rangle = \mathrm{E}(XY)$.

(Strictly, equivalence classes of a.s. equal r.v.s)

Projection theorem

Example: **Linear regression**

Given $y = (y_1, y_2, \ldots, y_n)' \in \mathbb{R}^n$, and $Z = (z_1, \ldots, z_q) \in \mathbb{R}^{n \times q}$, choose $\beta = (\beta_1, \ldots, \beta_q)' \in \mathbb{R}^q$ to minimize $\|y - Z\beta\|^2$.

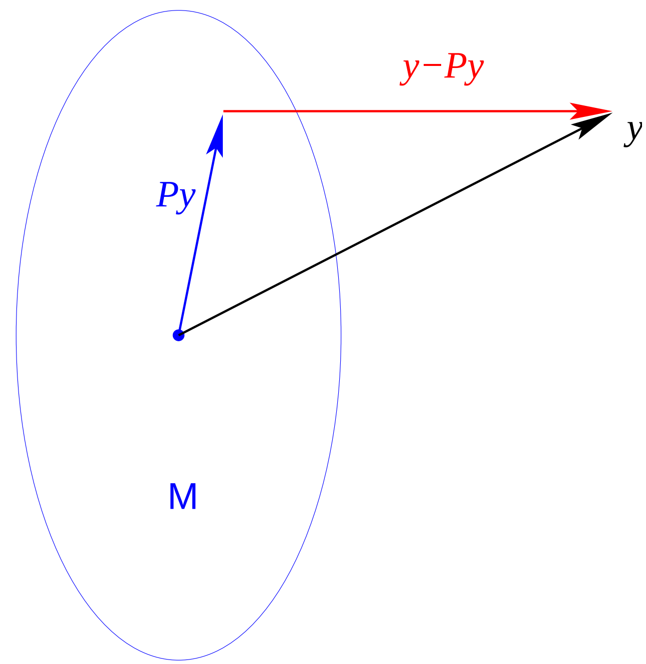
Here, $\mathcal{H} = \mathbb{R}^n$, with $\langle a, b \rangle = \sum_i a_i b_i$, and
$\mathcal{M} = \{Z\beta : \beta \in \mathbb{R}^q\} = \bar{\text{sp}}\,\{z_1, \ldots, z_q\}$.

Projection theorem

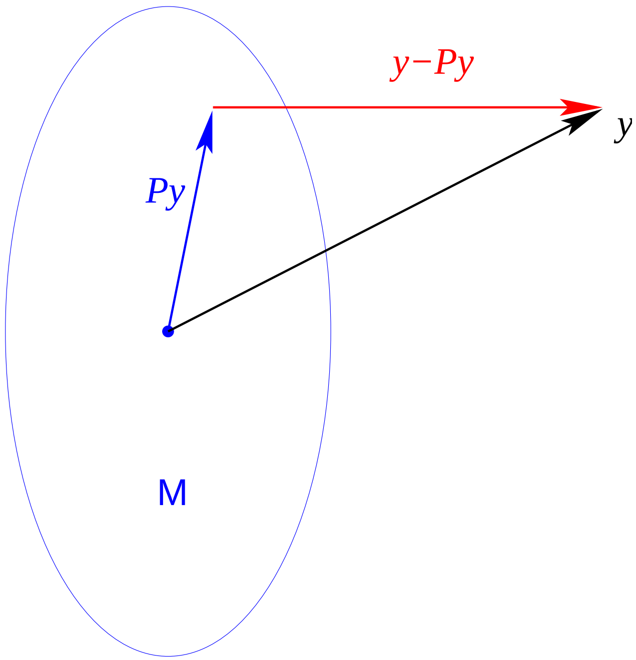
If $\mathcal{H}$ is a Hilbert space,
$\mathcal{M}$ is a closed subspace of $\mathcal{H}$,
and $y \in \mathcal{H}$,
then there is a point $Py \in \mathcal{M}$
(the **projection of** y **on** $\mathcal{M}$)
satisfying

1. $\|Py - y\| \leq \|w - y\|$
2. $\langle y - Py, w \rangle = 0$

for $w \in \mathcal{M}$.



Projection theorem



$$\langle y - Py, w \rangle = 0$$

$$\Leftrightarrow \quad \langle y - Z\hat{\beta}, z_i \rangle = 0, \quad \forall i$$

$$\Leftrightarrow \quad Z'Z\hat{\beta} = Z'y$$

$$\Leftrightarrow \quad \hat{\beta} = (Z'Z)^{-1}Z'y$$

"normal equations."

Projection theorem

Example: **Linear prediction**

Given $1, X_1, X_2, \ldots, X_n \in \{\text{mean 0 r.v.s } X : \mathrm{E}X^2 < \infty\}$,

choose $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{R}$
so that $Z = \alpha_0 + \sum_{i=1}^n \alpha_i X_i$ minimizes $\mathrm{E}(X_{n+m} - Z)^2$.

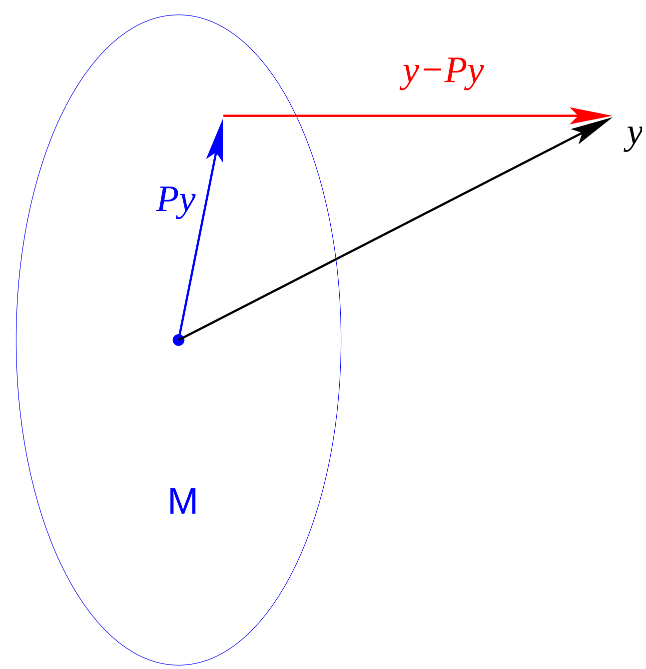
Here, $\langle X, Y \rangle = \mathrm{E}(XY)$,
$\mathcal{M} = \{Z = \alpha_0 + \sum_{i=1}^n \alpha_i X_i : \alpha_i \in \mathbb{R}\} = \bar{\mathrm{sp}}\{1, X_1, \ldots, X_n\}$, and
$y = X_{n+m}$.

Projection theorem

If $\mathcal{H}$ is a Hilbert space,
$\mathcal{M}$ is a closed subspace of $\mathcal{H}$,
and $y \in \mathcal{H}$,
then there is a point $Py \in \mathcal{M}$
(the **projection of** y **on** $\mathcal{M}$)
satisfying

1. $\|Py - y\| \leq \|w - y\|$
2. $\langle y - Py, w \rangle = 0$

for $w \in \mathcal{M}$.



Projection theorem: Linear prediction

Let X^n_{n+m} denote the best linear predictor:

$$\|X^n_{n+m} - X_{n+m}\|^2 \le \|Z - X_{n+m}\|^2 \quad \text{for all } Z \in \mathcal{M}.$$

The projection theorem implies the orthogonality

$$\begin{aligned} & \langle X^n_{n+m} - X_{n+m}, Z\rangle = 0 \quad \text{for all } Z \in \mathcal{M} \\ \Leftrightarrow \quad & \langle X^n_{n+m} - X_{n+m}, Z\rangle = 0 \quad \text{for all } Z \in \{1, X_1, \ldots, X_n\} \\ \Leftrightarrow \quad & \mathrm{E}\left(X^n_{n+m} - X_{n+m}\right) = 0 \\ & \mathrm{E}\left[\left(X^n_{n+m} - X_{n+m}\right) X_i\right] = 0 \end{aligned}$$

That is, the *prediction errors* ($X^n_{n+m} - X_{n+m}$) *are uncorrelated with the prediction variables* ($1, X_1, \ldots, X_n$).

Linear prediction

Error $(X_{n+m}^n - X_{n+m})$ is uncorrelated with the prediction variable 1:

$$\begin{aligned} & \mathrm{E}\left(X_{n+m}^n - X_{n+m}\right) = 0 \\ \Leftrightarrow \quad & \mathrm{E}\left(\alpha_0 + \sum_i \alpha_i X_i - X_{n+m}\right) = 0 \\ \Leftrightarrow \quad & \mu\left(1 - \sum_i \alpha_i\right) = \alpha_0. \end{aligned}$$

So $\quad X_{n+m}^n = \alpha_0 + \sum_i \alpha_i X_i \quad \Leftrightarrow \quad X_{n+m}^n - \mu = \sum_i \alpha_i \left(X_i - \mu\right).$

Thus, for forecasting, we can assume $\mu = 0$. So we'll ignore α_0.

One-step-ahead linear prediction

Write $X_{n+1}^n = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \cdots + \phi_{nn} X_1$

Prediction equations: $\mathrm{E}\left((X_{n+1}^n - X_{n+1}) X_i\right) = 0, \text{ for } i = 1, \ldots, n$

$$\Leftrightarrow \quad \sum_{j=1}^n \phi_{nj} \mathrm{E}\left(X_{n+1-j} X_i\right) = \mathrm{E}(X_{n+1} X_i)$$

$$\Leftrightarrow \quad \sum_{j=1}^n \phi_{nj} \gamma(i-j) = \gamma(i)$$

$$\Leftrightarrow \quad \Gamma_n \phi_n = \gamma_n,$$

One-step-ahead linear prediction

Prediction equations: $\Gamma_n \phi_n = \gamma_n.$

$$\Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & & \gamma(n-2) \\ \vdots & & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix},$$

$$\phi_n = (\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn})', \quad \gamma_n = (\gamma(1), \gamma(2), \ldots, \gamma(n))'.$$

Mean squared error of one-step-ahead linear prediction

$$\begin{aligned} P_{n+1}^n &= \mathrm{E}\left(X_{n+1} - X_{n+1}^n\right)^2 \\ &= \mathrm{E}\left(\left(X_{n+1} - X_{n+1}^n\right)\left(X_{n+1} - X_{n+1}^n\right)\right) \\ &= \mathrm{E}\left(X_{n+1}\left(X_{n+1} - X_{n+1}^n\right)\right) \\ &= \gamma(0) - \mathrm{E}\left(\phi_n' X X_{n+1}\right) \\ &= \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n, \end{aligned}$$

where $X = (X_n, X_{n-1}, \ldots, X_1)'$.

Mean squared error of one-step-ahead linear prediction

Variance is reduced:

$$\begin{aligned} P_{n+1}^{n} &= \mathrm{E}\left(X_{n+1} - X_{n+1}^{n}\right)^2 \\ &= \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n \\ &= \mathrm{Var}(X_{n+1}) - \mathrm{Cov}(X_{n+1}, X)\mathrm{Cov}(X, X)^{-1}\mathrm{Cov}(X, X_{n+1}) \\ &= \mathrm{E}\left(X_{n+1} - 0\right)^2 - \mathrm{Cov}(X_{n+1}, X)\mathrm{Cov}(X, X)^{-1}\mathrm{Cov}(X, X_{n+1}), \end{aligned}$$

where $X = (X_n, X_{n-1}, \ldots, X_1)'$.

Backcasting: Predicting m steps in the past

Given $X_1, \ldots, X_n$, we wish to predict X_{1-m} for $m > 0$.
That is, we choose $Z \in \mathcal{M} = \bar{\text{sp}}\{X_1, \ldots, X_n\}$ to minimize $\|Z - X_{1-m}\|^2$.

The prediction equations are

$$\begin{aligned} & \langle X_{1-m}^n - X_{1-m}, Z \rangle = 0 \quad \text{for all } Z \in \mathcal{M} \\ \Leftrightarrow \qquad & \mathrm{E}\left(\left(X_{1-m}^n - X_{1-m}\right) X_i\right) = 0 \quad \text{for } i = 1, \ldots, n. \end{aligned}$$

One-step backcasting

Write the least squares prediction of X_0 given $X_1, \ldots, X_n$ as

$$X_0^n = \phi_{n1} X_1 + \phi_{n2} X_2 + \cdots + \phi_{nn} X_n = \phi_n' X,$$

where the predictor vector is reversed: now $X = (X_1, \ldots, X_n)'$.
The prediction equations are

$$\begin{aligned}
& \mathrm{E}\left(\left(X_0^n - X_0\right) X_i\right) = 0 \quad \text{for } i = 1, \ldots, n \\
\Leftrightarrow \quad & \mathrm{E}\left(\left(\sum_{j=1}^{n} \phi_{nj} X_j - X_0\right) X_i\right) = 0 \\
\Leftrightarrow \quad & \sum_{j=1}^{n} \phi_{nj} \gamma(j - i) = \gamma(i) \\
\Leftrightarrow \quad & \Gamma_n \phi_n = \gamma_n.
\end{aligned}$$

One-step backcasting

The prediction equations are

$$\Gamma_n \phi_n = \gamma_n,$$

which is exactly the same as for forecasting, but with the indices of the predictor vector reversed: $X = (X_1, \ldots, X_n)'$ versus $X = (X_n, \ldots, X_1)'$.

Example: Forecasting AR(1)

AR(1) model: $X_t = \phi_1 X_{t-1} + W_t$

linear prediction of X_2: $X_2^1 = \phi_{11} X_1$

Prediction equation:

$$\begin{aligned} \gamma(0)\phi_{11} &= \gamma(1) \\ &= \text{Cov}(X_0, X_1) \\ &= \phi_1 \gamma(0) \end{aligned}$$

$$\Leftrightarrow \quad \phi_{11} = \phi_1.$$

Example: Backcasting AR(1)

AR(1) model: $X_t = \phi_1 X_{t-1} + W_t$

linear prediction of X_0: $X_0^1 = \phi_{11} X_1$

Prediction equation:

$$\begin{aligned} \gamma(0)\phi_{11} &= \gamma(1) \\ &= \text{Cov}(X_0, X_1) \\ &= \phi_1 \gamma(0) \end{aligned}$$

$$\Leftrightarrow \qquad \phi_{11} = \phi_1.$$

The prediction operator

For random variables $Y, Z_1, \ldots, Z_n$, define the **best linear prediction of** Y **given** $Z = (Z_1, \ldots, Z_n)'$ as the operator $P(\cdot|Z)$ applied to Y:

$$P(Y|Z) = \mu_Y + \phi'(Z - \mu_Z)$$

with $$\Gamma\phi = \gamma,$$

where $$\gamma = \text{Cov}(Y, Z)$$ $$\Gamma = \text{Cov}(Z, Z).$$

Properties of the prediction operator

1. $\mathrm{E}(Y - P(Y|Z)) = 0$, $\mathrm{E}((Y - P(Y|Z))Z) = 0$.

2. $\mathrm{E}((Y - P(Y|Z))^2) = \mathrm{Var}(Y) - \phi'\gamma$.

3. $P(\alpha_1 Y_1 + \alpha_2 Y_2 + \alpha_0 | Z) = \alpha_0 + \alpha_1 P(Y_1|Z) + \alpha_2 P(Y_2|Z)$.

4. $P(Z_i|Z) = Z_i$.

5. $P(Y|Z) = \mathrm{E}Y$ if $\gamma = 0$.

Example: predicting m steps ahead

Write $$X^n_{n+m} = \phi^{(m)}_{n1} X_n + \phi^{(m)}_{n2} X_{n-1} + \cdots + \phi^{(m)}_{nn} X_1$$

$$\Gamma_n \phi^{(m)}_n = \gamma^{(m)}_n,$$

with $$\Gamma_n = \mathrm{Cov}(X, X),$$

$$\gamma^{(m)}_n = \mathrm{Cov}(X_{n+m}, X)$$

$$= (\gamma(m), \gamma(m+1), \ldots, \gamma(m+n-1))'.$$

Also, $$\mathrm{E}((X_{n+m} - X^n_{n+m})^2) = \gamma(0) - \phi^{(m)\prime} \gamma^{(m)}_n.$$