SF2943 Time Series Analysis: Lecture 6

April 21, 2022

In the previous lecture, we presented one way to compute the acvf for a causal ARMA(p, q) process by expressing the process as a linear process and applying the formula for acvf for a linear process. In this lecture, we will introduce another way to calculate the acvf for a ARMA(p, q) process via solving linear difference equations. Moreover, we will discuss forecasting stationary time series with the best linear predictor.

1 Acvf for ARMA(p, q) process

Recall that a causal ARMA(p, q) process $\{X_t\}$ satisfies

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

and

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

For any $k = 0, 1, \cdots$, if we multiply the both sides of the equation by X_{t-k} and take expectations on each side (and assume $\theta_0 = 1$), we find that

$$\begin{aligned}
&\gamma(k) - \phi_1\gamma(k-1) - \cdots - \phi_p\gamma(k-p) \\
&\quad = E[X_t X_{t-k}] - \phi_1 E[X_{t-1}X_{t-k}] - \cdots - \phi_p E[X_{t-p}X_{t-k}] \\
&\quad = E[Z_t X_{t-k}] + \theta_1 E[Z_{t-1}X_{t-k}] + \cdots + \theta_p E[Z_{t-p}X_{t-k}] \\
&\quad = E\left[(Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q})X_{t-k}\right] \\
&\quad = E\left[\sum_{\ell=0}^{q} \theta_\ell Z_{t-\ell} \sum_{j=0}^{\infty} \psi_j Z_{t-k-j}\right] = \sum_{j=0}^{\infty}\sum_{\ell=0}^{q} \theta_\ell \psi_j E\left[Z_{t-\ell}Z_{t-k-j}\right] \\
&\quad = \begin{cases} 0, & \text{for } k \geq q+1 \\ \sigma^2 \sum_{j=0}^{\infty} \theta_{k+j}\psi_j, & \text{for } 0 \leq k \leq q \end{cases}. \qquad (1.1)
\end{aligned}$$

The last equation holds since for any j and k, $E\left[Z_{t-\ell}Z_{t-k-j}\right]$ is not zero only when $\ell = k + j$. Moreover, when $\ell = k + j$, $E\left[Z_{t-\ell}Z_{t-k-j}\right] = E[Z_{t-k-j}^2] = \sigma^2$. However, $\ell \in \{0, 1, \ldots, q\}$, this implies that for any $k \geq q + 1$ and $j \in \mathbb{N} \cup \{0\}$, we have $k + j \neq \ell$.

Notice that if $k \geq p$, $\gamma(k) - \phi_1\gamma(k-1) - \cdots - \phi_p\gamma(k-p)$ contains p independent variables $\gamma(k), \ldots, \gamma(k-p)$. For any $k \geq m \doteq \max(p, q+1)$, we have

$$\gamma(k) - \phi_1\gamma(k-1) - \cdots - \phi_p\gamma(k-p) = 0, \text{ for any } k \geq m.$$

This set of homogeneous linear difference equations, according to linear algebra, has a unique solution of the following form

$$\gamma(h) = \alpha_1\xi_1^{-h} + \alpha_2\xi_2^{-h} + \cdots + \alpha_p\xi_p^{-h} \tag{1.2}$$

for any $h \geq m - p$, where $\xi_1, \ldots, \xi_p$ are the roots (assumed to be distinct) of the AR polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0$, and $\alpha_1, \ldots, \alpha_p$ are unknown constants to be determined.

To find $\alpha_1, \ldots, \alpha_p$ as well as $\gamma(h)$ with $0 \in \{0, \cdots, m-p-1\}$, we plug (1.2) into (1.1) and solve the linear equations.

Using this method, let's demonstrate how to calculate acvf by considering a concrete example.

Example 1.1 *Consider the causal ARMA$(1,1)$ process $\{X_t\}$, i.e., $p = 1$ and $q = 1$. Recall that*

$$X_t - \phi X_{t-1} = Z_t + \theta Z_t$$

and from our calculation in lecture 5

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

with

$$\psi_j = \begin{cases} 1, & \text{for } j = 0 \\ \phi^{j-1}(\theta + \phi), & \text{for } j > 0 \end{cases}.$$

This implies that (1.1) becomes (for $k = 0$)

$$\gamma(0) - \phi\gamma(1) = \gamma(0) - \phi\gamma(-1) = \sigma^2 \sum_{j=0}^{\infty} \theta_j\psi_j = \sigma^2(\theta_0\psi_0 + \theta_1\psi_1) = \sigma^2(1 + \theta(\theta + \phi)) \tag{1.3}$$

and (for $k = 1$)

$$\gamma(1) - \phi\gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \theta_{1+j}\psi_j = \sigma^2\theta_1\psi_0 = \sigma^2\theta \tag{1.4}$$

and

$$\gamma(k) - \phi\gamma(k-1) = 0$$

for any $k \geq 2$. Therefore, solving (1.3) and (1.4) gives

$$\gamma(0) = \sigma^2\left(1 + \frac{(\theta+\phi)^2}{1-\phi^2}\right) \text{ and } \gamma(1) = \sigma^2\left(\theta + \phi + \frac{(\theta+\phi)^2\phi}{1-\phi^2}\right)$$

and for any $h \geq 2$, $\gamma(h) = \phi^h\gamma(1)$. This is the same as we derived in the previous lecture (where we used the "formula" method).

Remark 1.1 *This remark is only my personal take on these two methods for calculating acvf. In my opinion, the formula method (as we mentioned in the previous lecture) is more straightforward and maybe a better way. The reason is that both methods require computing* $\{\psi_j\}_j$*, but it is the only crucial step for the first method. While for the second method, one still needs to solve linear equations as well as derive/remember equation* (1.1) *and/or* (1.2)*.*

2 Forecasting Stationary Time Series

We now consider the problem of predicting/forecasting the value X_{n+h} for some $h \in \mathbb{N}$ given $\{X_1, \cdots, X_n\}$ coming from a stationary time series. It is reasonable to consider a predictor of X_{n+h} using $f(X_1, \ldots, X_n)$ for some function $f : \mathbb{R}^n \to \mathbb{R}$. Also, a natural way to evaluate the performance of a predictor is by considering the mean squared error (MSE)

$$E\left[(X_{n+h} - f(X_1, \ldots, X_n))^2\right].$$

With this performance measure, we say that a predictor $f(X_1, \ldots, X_n)$ is the best if it minimizes the MSE within all functions $f : \mathbb{R}^n \to \mathbb{R}$. However, such minimization problem is infinite dimensional, thus it is extremely difficult to find the best predictor. Alternatively, we can consider finding the best linear predictor, i.e., consider

$$f(X_1, \ldots, X_n) = a_0 + a_1 X_n + \cdots + a_n X_1,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{R}$. Finding the best linear predictor is feasible since we only need to find the minimizer of

$$\min_{a_0, a_1, \ldots, a_n \in \mathbb{R}} E\left[(X_{n+h} - (a_0 + a_1 X_n + \cdots + a_n X_1))^2\right],$$

which is a $n+1$ dimensional problem. In fact, we can find the optimal $a_0, a_1, \ldots, a_n$ with the help of Calculus.

To be more precise, let

$$S(a_0, \ldots, a_n) = E\left[(X_{n+h} - (a_0 + a_1 X_n + \cdots + a_n X_1))^2\right]$$

and let μ be the mean and γ be the acvf for the time series $\{X_t\}$. For simplicity, we consider the case $\mu = 0$.

Remark 2.1 *For any stationary time series* $\{Y_t\}$ *with nonzero* μ*, we can consider* $X_t = Y_t - \mu$*, and fine the best linear predictor of* X_t *denoted by* $a_0 + a_1 X_n + \cdots + a_n X_1$*. Then it is easy to show that the best linear predictor of* Y_t *is* $\mu + a_0 + a_1(Y_n - \mu) + \cdots + a_n(Y_1 - \mu)$*.*

Then

$$\begin{aligned}
&S(a_0,\ldots,a_n)\\
&\quad= E\left[\left((X_{n+h}-(a_1X_n+\cdots+a_nX_1)\right)-a_0\right)^2\right]\\
&\quad= E\left[\left(X_{n+h}-(a_1X_n+\cdots+a_nX_1)\right)^2\right]-2a_0E\left[X_{n+h}-(a_1X_n+\cdots+a_nX_1)\right]+a_0^2\\
&\quad= E\left[\left(X_{n+h}-(a_1X_n+\cdots+a_nX_1)\right)^2\right]-2a_0\left(EX_{n+h}-a_1EX_n-\cdots-a_nEX_1\right)+a_0^2\\
&\quad= E\left[\left(X_{n+h}-(a_1X_n+\cdots+a_nX_1)\right)^2\right]+a_0^2.
\end{aligned}$$

By taking partial derivatives with respect to a_i for $i=0,1,\ldots,n$ separately yields

$$\frac{\partial S(a_0,\ldots,a_n)}{\partial a_0}=2a_0,$$

and for any $i=1,\ldots,n$

$$\begin{aligned}
\frac{\partial S(a_0,\ldots,a_n)}{\partial a_i}&=\frac{\partial}{\partial a_i}E\left[\left(X_{n+h}-(a_1X_n+\cdots+a_nX_1)\right)^2\right]\\
&=E\left[\frac{\partial}{\partial a_i}\left(X_{n+h}-(a_1X_n+\cdots+a_nX_1)\right)^2\right]\\
&=-2E\left[\left(X_{n+h}-(a_1X_n+\cdots+a_nX_1)\right)X_i\right]\\
&=-2\gamma(h+n-i)+2a_1\gamma(n-i)+\cdots+2a_n\gamma(1-i)\\
&=-2\left(\gamma(h+n-i)-\sum_{j=1}^{n}a_j\gamma(n-j+1-i)\right).
\end{aligned}$$

The optimal $a_0,a_1,\ldots,a_n$ satisfies

$$\frac{\partial S(a_0,\ldots,a_n)}{\partial a_i}=0 \text{ for all } i=0,1,\ldots,n,$$

which is equivalent to $a_0=0$ and

$$\gamma(h+n-i)-\sum_{j=1}^{n}a_j\gamma(n-j+1-i)=0 \text{ for all } i=0,1,\ldots,n. \tag{2.1}$$

Using a matrix notation, we can rewrite (2.1) in a more succinct way:

$$\Gamma_n\mathbf{a}_n=\gamma_n(h), \tag{2.2}$$

where $\mathbf{a}_n=(a_1,\ldots,a_n)'$, $\Gamma_n=[\gamma(i-j)]_{i,j=1}^n$, and $\gamma_n(h)=(\gamma(h),\gamma(h+1),\ldots,\gamma(h+n-1))'$.

To conclude, for a time series $\{X_t\}$ with mean 0, the best linear predictor of X_{n+h} given $X_1,\ldots,X_n$, denoted by P_nX_{n+h}, is

$$P_nX_{n+h}=a_0+a_1X_n+\cdots+a_nX_1,$$

with $a_0 = 0$ and $\mathbf{a}_n = (a_1, \ldots, a_n)'$ solves (2.2). In addition, for a time series $\{X_t\}$ with mean μ, the best linear predictor of X_{n+h} given $X_1, \ldots, X_n$, denoted by $P_n X_{n+h}$, is

$$P_n X_{n+h} = a_0 + a_1 X_n + \cdots + a_n X_1,$$

with

$$a_0 = \mu \left(1 - \sum_{i=1}^{n} a_i\right)$$

and $\mathbf{a}_n = (a_1, \ldots, a_n)'$ solves (2.2).

Properties of $P_n X_{n+h}$:

1. $P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)$, where $\mathbf{a}_n = (a_1, \ldots, a_n)'$ satisfies (2.2).
2. $E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \mathbf{a}_n' \gamma_n(h)$, where $\gamma_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))'$.
3. $E(X_{n+h} - P_n X_{n+h}) = 0$. Namely, $P_n X_{n+h}$ is an unbiased estimator of X_{n+h}.
4. $E[(X_{n+h} - P_n X_{n+h}) X_j] = 0$ for $j = 1, \ldots, n$.

Property 4 says that the expectation of "Error" $(X_{n+h} - P_n X_{n+h})$ times "Predictor variable" (X_j) is always zero.

2.1 One-Step prediction $P_n X_{n+1}$

Let's try to find the best linear predictor for a concrete example.

Example 2.1 (One-Step Prediction of an AR(1) Process) *Consider an AR(1) process $\{X_t\}$. We know that X_t satisfies*

$$X_t = \phi X_{t-1} + Z_t,$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $|\phi| < 1$.

From our previous lectures, we know that for any $h \in \mathbb{Z}$,

$$\gamma(h) = \phi^{|h|} \frac{\sigma^2}{1 - \phi^2}$$

and we know that $EX_t = 0$. Therefore, $a_0 = 0$,

$$\Gamma_n = [\gamma(i-j)]_{i,j=1}^{n} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(-1) & \gamma(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(-n+1) & \cdots & \cdots & \gamma(0) \end{pmatrix} = \frac{\sigma^2}{1 - \phi^2} \begin{pmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \cdots & \cdots & 1 \end{pmatrix}$$

and

$$\gamma_n(1) = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(h) \end{pmatrix} = \frac{\sigma^2}{1 - \phi^2} \begin{pmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{pmatrix}.$$

Moreover, (2.2) becomes

$$\frac{\sigma^2}{1-\phi^2}\begin{pmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \cdots & \cdots & 1 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \frac{\sigma^2}{1-\phi^2}\begin{pmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{pmatrix},$$

which gives $\mathbf{a}_n = (a_1, \ldots, a_n)' = (\phi, 0, \ldots, 0)'$. Hence, the best linear predictor of X_{n+1} in terms of $\{X_1, \ldots, X_n\}$ is

$$P_n X_{n+1} = \mathbf{a}_n' \mathbf{X}_n = \phi X_n$$

with mean squared error

$$E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \mathbf{a}_n' \gamma_n(h) = \frac{\sigma^2}{1-\phi^2} - \phi\gamma(1) = \sigma^2.$$

Remark 2.2 *Intuitively speaking, we know that for an AR(1) process, $X_t = \phi X_{t-1} + Z_t$ for any t, where Z_t is a mean zero external noise and uncorrelated to X_{t-1}. Hence, given all the information $\{X_1, \ldots, X_n\}$, in order to predict X_{n+1}, the information on X_n should be sufficient, and we would expect that ϕX_n provides a reasonable prediction of X_{n+1} (the best prediction of Z_n is zero since it is mean zero). This predictor derived from this intuition, ϕX_n, turns out to be the same as the best linear predictor we just found.*

2.2 The Durbin-Levinson Algorithm

So far we have shown that for any fixed $n \in \mathbb{N}$, the way to find the best linear predictor of X_{n+1} in terms of $\{X_1, \ldots, X_n\}$ is by solving $\phi_n = (\phi_{n,1}, \phi_{n,2}, \ldots, \phi_{n,n})'$ from

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

where $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n$, and $\gamma_n = (\gamma(1), \gamma(2), \ldots, \gamma(n))'$. When we obtain a new data X_{n+1}, we can certainly find the best linear predictor of X_{n+2} in terms of $\{X_1, \ldots, X_n, X_{n+1}\}$ is by solving $\phi_{n+1} = (\phi_{n+1,1}, \phi_{n+1,2}, \ldots, \phi_{n+1,n+1})'$ from

$$\Gamma_{n+1} \phi_{n+1} = \gamma_{n+1}(1),$$

where $\Gamma_{n+1} = [\gamma(i-j)]_{i,j=1}^{n+1}$, and $\gamma_{n+1} = (\gamma(1), \gamma(2), \ldots, \gamma(n), \gamma(n+1))'$. However, this might be inefficient since it is obvious that Γ_n and Γ_{n+1} share lots of similarities, so do γ_n and γ_{n+1}. It turns out that in some cases, we do not need to solve the whole linear system again, and instead we can use $P_n X_{n+1}$ to compute $P_{n+1} X_{n+2}$

Suppose Γ_n is a non-singular/invertible matrix for every $n \in \mathbb{N}$. Then Γ_n is invertible, we find $\phi_n = \Gamma_n^{-1} \gamma_n$ and the corresponding mean squared error is

$$\nu_n \doteq E(X_{n+1} - P_n X_{n+1})^2 = \gamma(0) - \phi_n' \gamma_n.$$

Remark 2.3 *A sufficient condition for nonsingularity of all the autocovariance matrices $\Gamma_1, \Gamma_2, \ldots$ is $\gamma(0) > 0$ and $\gamma(h) \to 0$ as $h \to \infty$. (See Proposition 5.1.1 in Time Series: Theory and Methods for a proof.)*

Theorem 2.4 (The Durbin-Levinson Algorithm) *$\phi_{1,1} = \gamma(1)/\gamma(0)$ and $\nu_0 = \gamma(0)$. Moreover, for any $n \in \mathbb{N}$, the best linear predictor of X_{n+1} in terms of $X_1, \ldots, X_n$ is*

$$P_n X_{n+1} = \phi_{n,1} X_n + \cdots + \phi_{n,n} X_n,$$

where the coefficients $\phi_{n,1}, \ldots, \phi_{n,n}$ and ν_n can be computed recursively from $\phi_{n-1,1}, \ldots, \phi_{n-1,n-1}$, ν_{n-1}, and the equations

$$\phi_{n,n} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] \nu_{n-1}^{-1},$$

$$\begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{n,n} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}$$

and

$$\nu_n = \nu_{n-1}[1 - \phi_{n,n}^2].$$

Proof. See the book. ■

3 Prediction of Second-Order Random Variables

Our theory on the prediction of stationary time series can be easily extended to the prediction of second-order random variables. Let Y and $W_1, , W_n$ be random variables with finite second moments and assume that the means $\mu = EY$, $\mu_i = EW_i$ and covariances $\mathrm{Cov}(Y, Y)$, $\mathrm{Cov}(Y, W_i)$, and $\mathrm{Cov}(W_i, W_j)$ for any $i, j \in \{1, \ldots, n\}$ are all known. Denote the random vector $\mathbf{W} = (W_n, \ldots, W_1)'$, the vector of means $\boldsymbol{\mu}_W = (\mu_n, \ldots, \mu_1)'$, the vector of covariances

$$\gamma = \mathrm{Cov}(Y, \mathbf{W}) \doteq (\mathrm{Cov}(Y, W_i), \mathrm{Cov}(Y, W_i), \ldots, \mathrm{Cov}(Y, W_i))',$$

and the covariance matrix

$$\Gamma = \mathrm{Cov}(\mathbf{W}, \mathbf{W}) \doteq [\mathrm{Cov}(W_i, W_j)]_{i,j=1}^{n}.$$

Then by the same arguments used in the calculation of $P_n X_{n+h}$, the best linear predictor of Y in terms of $\{1, W_n, \ldots, W_1\}$ is

$$P(Y|\mathbf{W}) = \mu_Y + \mathbf{a}'(\mathbf{W} - \boldsymbol{\mu}_W),$$

where $\mathbf{a} = (a_1, \ldots, a_n)'$ is any solution of

$$\Gamma \mathbf{a} = \gamma.$$

The mean squared error of the predictor is

$$E\left[(Y - P(Y|\mathbf{W}))^2\right] = \mathrm{Var}(Y) - \mathbf{a}'\gamma.$$

Remark 3.1 *The main difference between general second-order random variables and a stationary time series is that the former can have different means* $\mathbf{W} = (W_n, \ldots, W_1)'$*, but the latter has the same mean* $\mu_1 = \cdots = \mu_n = \mu$*.*

$P(\cdot|\mathbf{W})$ has many analogous properties as $P_n X_{n+h}$. We state some important properties of $P(\cdot|\mathbf{W})$ in the following.

Properties of the Prediction Operator $P(\cdot|\mathbf{W})$:
Suppose $EU^2 < \infty$, $EV^2 < \infty$, $\Gamma = \mathrm{Cov}(\mathbf{W}, \mathbf{W})$, and $\beta, \alpha_1, \ldots, \alpha_n$ are constant.

1. $P(U|\mathbf{W}) = EU + \mathbf{a}'(\mathbf{W} - E\mathbf{W})$, where $\Gamma\mathbf{a} = \mathrm{Cov}(U, \mathbf{W})$.
2. $E[(U - P(U|\mathbf{W}))\mathbf{W}] = \mathbf{0}$ and $E[U - P(U|\mathbf{W})] = 0$.
3. $E[(U - P(U|\mathbf{W}))^2] = \mathrm{Var}(U) - \mathbf{a}'\mathrm{Cov}(U, \mathbf{W})$.
4. $P(\alpha_1 U + \alpha_2 V + \beta|\mathbf{W}) = \alpha_1 P(U|\mathbf{W}) + \alpha_2 P(V|\mathbf{W}) + \beta$.
5. $P(U|\mathbf{W}) = EU$ if $\mathrm{Cov}(U, \mathbf{W}) = \mathbf{0}$.
6. If $U = g(\mathbf{W})$ for some function $g : \mathbb{R}^n \to \mathbb{R}$, then $P(U|\mathbf{W}) = U$.

The following example shows that these properties could help identify the best linear predictor.

Example 3.1 (One-Step Prediction of an AR(p) Series) *Consider an AR(p) process* $\{X_t\}$*, i.e.,* X_t *satisfies*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t.$$

Recall that Z_t *is uncorrelated with* X_s *for each* $s < t$ *and* $EZ_t = 0$ *for each* t*.*

Then if $n > p$, let $\mathbf{X} = (X_n, \ldots, X_1)'$, we can apply the prediction operator P_n to both sides of the equation above with $t = n + 1$ to find that the best linear predictor is

$$\begin{aligned} P_n X_{n+1} &= P(X_{n+1}|\mathbf{X}) = P(\phi_1 X_n + \cdots + \phi_p X_{n+1-p} + Z_{n+1}|\mathbf{X}) \\ &= \phi_1 P(X_1|\mathbf{X}) + \cdots + \phi_n P(X_n|\mathbf{X}) + P(Z_{n+1}|\mathbf{X}) \\ &= \phi_1 X_n + \cdots + \phi_p X_{n+1-p}. \end{aligned}$$

The last equation holds since we use property 5 to find $P(Z_{n+1}|\mathbf{X}) = EZ_{n+1} = 0$ and property 6 to find $P(X_i|\mathbf{X}) = X_i$ for $i = n + 1 - p, \ldots, n$.