

Stat 443: Notation and big picture for forecasting

The notation and some probability-based expressions in this document might look difficult on first and second pass. Return to review this document periodically over the term after seeing how the notation is used to explain forecasts.

1 Notation

Note: the reference book by Hyndman and Athanasopoulos does not have good notation to distinguish random variables in stochastic models and realized values in data sets. The proper notation is needed to show derivations and introduce techniques for mathematical analysis of time series models (a) that are presented in this course and (b) that you might encounter in further reading and future work.

t is the time index, T is the length of time series.

Otherwise mostly commonly:

- Upper case roman letters for random variables,
- lower case roman letters for realized data values,
- lower case Greek letters for parameters,
- exception of Greek letter ϵ to denote a random variable that represents deviation from the mean (or innovation),
- overhead hat or caret $\hat{}$ for estimated quantity,

Observed data $y_1, \dots, y_t, \dots, y_T$ of a variable y at time points $1, \dots, T$. The time unit is usually day, week, month, quarter, year. Sometimes there are vectors of exogenous variables or covariates $\mathbf{x}_1, \dots, \mathbf{x}_T$. For example, y =monthly sales at a clothing store, \mathbf{x} =(average temperature, unemployment rate, amount of advertising, ...)

Quantity	Stat443	H&A
time index	t	t
length of time series	T	T
observed value	y_t	y_t
random variable	Y_t	y_t
1-step forecast for time $t + 1$ using data to time t	$\hat{y}_{t+1 t}$	$\hat{y}_{t+1 t}$
h -step forecast for time $t + h$ using data to time t	$\hat{y}_{t+h t}$	$\hat{y}_{t+h t}$
innovation (rv at time t indep. of past)	ϵ_t	ϵ_t
forecast error	$e_t(h), E_t(h)$	$e_t(h)$
exogenous variable vector	\mathbf{x}_t	\mathbf{x}_t
difference of consecutive lag	y'_t or $Y'_t?$	y'_t
level (exponential smoothing) $a[t]$ in R	ℓ_t	ℓ_t
slope or trend $b[t]$ in R	b_t	b_t
seasonal $s[t]$ in R	s_t	s_t
past data at time $t + 1$	y_1, \dots, y_t or \mathcal{F}_t	y_1, \dots, y_t
autocorrelation lag k	ρ_k	$?$
sample autocorrelation lag k	$\hat{\rho}_k$	r_k
partial autocorrelation lag k	α_k	$?$
sample partial autocorrelation lag k	$\hat{\alpha}_k$	$?$

Quantities inside $E, \text{Var}, \text{Cov}, \text{Cor}$ that are random variables should be represented in upper case.

The lower case letter f , possibly with subscripts, is often used in statistics and probability for a probability density function or probability mass function. The upper case letter F , possibly with subscripts, is used for the cumulative distribution function.

Conditional expectations and variances are used a lot in time series forecasting because (a) the forecasts are based on the expected value of the next observation given the observed past, (b) prediction intervals for forecasts make use of the variance of the next observation given the observed past.

Let Y_1, Y_2, \dots, Y_n be continuous random variables (for Stat 443 Y is continuous and not discrete; for theory for discrete time series, there are many more probabilistic concepts that are needed). The conditional density of Y_n given $Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}$ is denoted as

$$f_{Y_n|Y_1, \dots, Y_{n-1}}(\cdot | y_1, \dots, y_{n-1}),$$

With $Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}$ given, this conditional density is a mapping from $(-\infty, \infty)$ to $[0, \infty)$. From a conditional density, conditional probabilities, expectation and variance

can be defined.

$$\begin{aligned}\Pr(a < Y_n \leq b | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) &= \int_a^b f_{Y_n|Y_1, \dots, Y_{n-1}}(z | y_1, \dots, y_{n-1}) dz \\ E(Y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) &= \mu_{Y_n}(y_1, \dots, y_{n-1}) = \int_{-\infty}^{\infty} z f_{Y_n|Y_1, \dots, Y_{n-1}}(z | y_1, \dots, y_{n-1}) dz \\ \text{Var}(Y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) &= \sigma_{Y_n}^2(y_1, \dots, y_{n-1}) \\ &= \int_{-\infty}^{\infty} [z - \mu_{Y_n}(y_1, \dots, y_{n-1})]^2 f_{Y_n|Y_1, \dots, Y_{n-1}}(z | y_1, \dots, y_{n-1}) dz\end{aligned}$$

Stat 443 mainly introduces the class of linear Gaussian time series for which integration is not necessary when using **rules for expectations, variances and covariances of linear combination of random variables**.

That is, Gaussian time series models have stochastic representations so that integration might not be needed.

For example, consider the stochastic representation

$$[Y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}] \sim g_n(y_1, \dots, y_{n-1}) + \epsilon_n,$$

where g_n is a continuous function of past observations. The meaning of this notation in words is: “The conditional distribution of Y_n given $Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}$ is stochastically the same as $g_n(y_1, \dots, y_{n-1}) + \epsilon_n$ where ϵ_n is a random variable.

In a time series context, Y_n is a future observation, y_1, \dots, y_{n-1} are realized values of past random variables, and ϵ_n is an innovation random variable (usually with mean 0).

Understanding the proper use of notation with upper and lower case is important for interpretation on conditional quantities. What if only Y_i ’s or only y_i ’s occur after the vertical bar?

- $E(Y_n | y_1, \dots, y_{n-1})$ can be shorthand for $E(Y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1})$ if the context of the conditioning random variables is clear, e.g., the meaning of $E(Y_n | 0.5, 1.3, 0.9, \dots, 4.2)$ must be clear. For general multivariate densities:

$$\begin{aligned}\mu_{Y_n|Y_1, \dots, Y_{n-1}}(y_1, \dots, y_{n-1}) &= \mu_{Y_n|\mathcal{F}_{n-1}}(y_1, \dots, y_{n-1}) \\ &:= E(Y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) \\ &= \int_{-\infty}^{\infty} z f_{Y_n|Y_1, \dots, Y_{n-1}}(z | y_1, \dots, y_{n-1}) dz,\end{aligned}$$

where $f_{Y_n|Y_1, \dots, Y_{n-1}}$ is a conditional density. The conditional mean depends on the conditioning values y_1, \dots, y_{n-1} .

- $E(Y_n|Y_1, \dots, Y_{n-1})$ is shorthand for $\mu_{Y_n|\mathcal{F}_{n-1}}(Y_1, \dots, Y_{n-1})$, and it is a random variable that is a function of Y_1, \dots, Y_{n-1} , from the preceding item.
- $E(y_n|y_1, \dots, y_{n-1}) = y_n$ because y_n is a constant.

2 Big picture for time series modelling and forecasting

The data series y_1, \dots, y_T is a realization of the random sequence Y_1, \dots, Y_T . A general stochastic model is:

$$[Y_t|Y_1 = y_1, \dots, Y_{t-1} = y_{t-1}] \stackrel{d}{=} g_t(y_1, \dots, y_{t-1}; \boldsymbol{\theta}) + \epsilon_t.$$

$\boldsymbol{\theta}$ is a parameter to be estimated based on training set y_1, \dots, y_n , where $T/2 < n < T$. Suppose the estimate is $\hat{\boldsymbol{\theta}}$. If $E[\epsilon_t] = 0$ for all t , then

$$g_t(y_1, \dots, y_{t-1}; \hat{\boldsymbol{\theta}})$$

is the fitted value for $t = 2, \dots, n$, and it is the 1-step forecast value for $t = n+1, \dots, T$. In practice, for an additive innovation, there are different models/methods for the g_t functions (denoted as $g_t^{(m)}$ for method m).

Exercise to check your understanding from previous courses: What is a 50% prediction or 1-step forecast interval for Y_t for $t \in \{n+1, \dots, T\}$? Assume that the ϵ_t are independent and identically distributed (iid) with mean 0 and variance σ_ϵ^2 , and ...(?).

Let the 1-step forecast at time t for method m be denoted as $\hat{y}_{t+1|t}^{(m)}$.

A common criterion for comparison is based on the root mean square prediction error (RMSPE):

$$\left\{ (T - n) \sum_{t=n+1}^T (\hat{y}_{t+1|t}^{(m)} - y_t)^2 \right\}^{1/2}$$

A better forecasting method has a smaller RMSPE.

Exercise to check your understanding: What are unstated assumptions in order that this RMSPE is meaningful.

More sophisticated methods to compare forecasts (beyond the scope of this course) are based on 50%, 60%, 70%, 80%, 90% forecast intervals.

Inferential statistics goes beyond *point forecasts* to *interval forecasts*. Compare point estimate and confidence or prediction intervals in regression.