Figure 22.6 (a) The relation between $g_s[k]$ and $f_s[k]$; (b) An optimal predictor $f_s[k] = g_s[k + p]$ that acts on white noise.

Since the white noise $W_t$ can be created by passing the signal $S_t$ into a filter $1/G_s(z)$

$$H_{opt}(z) = \frac{1}{G_s(z)} [G_s(z)z^p]_+,$$  \hspace{1cm} (22.117)
Figure 22.7 An optimal predictor $H_{\text{opt}}(z)$ that acts on the input signal $S_t$.

$$H_{\text{opt}}(z) = \frac{F_s(z)}{G_s(z)}$$

$$h_{\text{opt}}[k] = Z^{-1} \left\{ \frac{F_s(z)}{G_s(z)} \right\}.$$ 

$$\sum_{k=0}^{\infty} |g_s[k]|^2 = E \left[ |S_{t+p}|^2 \right] = \|S_{t+p}\|^2 \quad (22.119)$$

$$\sum_{k=0}^{\infty} |f_s[k]|^2 = \sum_{k=p}^{\infty} |g_s[k]|^2 = E \left[ |\hat{S}_{t+p}|^2 \right] = \|\hat{S}_{t+p}\|^2 \quad (22.120)$$
\[ \|e_{t+p}\|^2 = \|S_{t+p}\|^2 - \|\hat{S}_{t+p}\|^2 = \sum_{k=0}^{p-1} |g_s[k]|^2 \]  

(22.121)

**Example 22.3: Pure predictor of an autoregressive signal**

\[ X_t = S_t \]

\[ P_{ss}(z) = \frac{|A|^2}{(1 - \alpha z^{-1})(1 - \overline{\alpha} z)} \]

\[ P_s(\omega) = \frac{|A|^2}{2\pi (1 + |\alpha|^2 - 2\Re\{\alpha\} \cos \omega)} \]  

(22.122)

\[ G_s(z) = \frac{A}{1 - \alpha z^{-1}} \]

Then, by referring to Figure 22.5, the filtered noise \( \{Z_t\} \) and the white noise \( \{W_t\} \) are related by

\[ Z_t = \alpha Z_{t-1} + AW_t \]  

(22.123)

an *autoregressive sequence* of first-order denoted as AR(1)
If the white noise $W_t$ is Gaussian, $Z_t$ is a Gauss-Markov process $S_t$, which is statistically equivalent to $Z_t$, is also a GMP.

$$g_s[k] = \begin{cases} 0, & k < 0, \\ A \alpha^k, & k = 0, 1, 2, \ldots \end{cases} \tag{22.124}$$

$$f_s[k] = \begin{cases} 0, & k < 0, \\ A \alpha^{k+p}, & k \geq 0. \end{cases}$$

$$F_s(z) = \frac{A \alpha^p}{1 - \alpha z^{-1}}.$$  

$$H_{opt}(z) = \frac{F_s(z)}{G_s(z)} = \alpha^p. \tag{22.125}$$

$$\hat{S}_{t+p} = \alpha^p S_t.$$
The dashed curve is the output of a pure Wiener predictor, which estimates $S_t$ based on $S_{t-2}$.
The predicted waveform is a scaled (by $\alpha^p = 2^{-1/2}$) and shifted (by $p = 2$ time units) version of the actual signal waveform

$$E[S_{t+p} | S_t] = \rho S_t = \alpha^p S_t, \quad p \geq 0.$$ 

$$\text{Var}[S_{t+p} | S_t] = \sigma_s^2 (1 - \rho^2) = \sigma_s^2 (1 - \alpha^{2p})$$

$$|Z_t|^2 = |\alpha Z_{t-1}|^2 + |AW_t|^2 + 2\Re\{\alpha \overline{A} Z_{t-1} \overline{W}_t\}$$

$$\sigma_s^2 = |\alpha|^2 \sigma_s^2 + |A|^2.$$ 

$$\sigma_s^2 = \frac{|A|^2}{1 - |\alpha|^2} \quad (22.128)$$
Theorem 22.4 (Optimality of the best linear pure predictor for a Gaussian process).

If the signal $S_t$ is a Gaussian process, the best linear pure predictor $\hat{S}_{t+p}$ is as good as any predictor, linear or nonlinear, in the minimum mean square error (MMSE) sense.

Proof: Let $Y_t$ be any prediction of $S_{t+p}$ based on the input prior to time $t$, i.e., $\{S_{t'}, t' < t\}$. Then $Y_t$ is independent of $\{W_k, k > t\}$.

Since
$$Z_t = Y_t - \hat{S}_{t+p} = Y_t - \sum_{k=-\infty}^{t} g_s[t + p - k]W_k \quad (22.130)$$

$Z_t$ is also independent of $\{W_k, k > t\}$. Furthermore,

$$S_{t+p} - \hat{S}_{t+p} = \sum_{k=t+1}^{t+p} g_s[t + p - k]W_k$$

$Z_t$ and $S_{t+p} - \hat{S}_{t+p}$ are independent, hence orthogonal.
\[ E \left[ |S_{t+p} - Y_t|^2 \right] = E \left[ |S_{t+p} - \hat{S}_{t+p} - Z_t|^2 \right] \]
\[ = E \left[ |S_{t+p} - \hat{S}_{t+p}|^2 \right] + E \left[ |Z_t|^2 \right] \]
\[ \geq E \left[ |S_{t+p} - \hat{S}_{t+p}|^2 \right] \]  \hspace{1cm} (22.131)

Hence, the mean square error of any prediction \( Y_t \) cannot be made smaller than the error of the best linear prediction \( \hat{S}_{t+p} \).

The above theorem states that if \( S_t \) is a Gaussian process, then \( E[S_{t+p}|S_k, \ u \leq t] = \sum_{k=0}^{\infty} h[k] S_{t-k} \) with some linear predictor \( h[k], \ t \geq 0 \).
The first step in solving the general case is to factor out $P_{xx}(z)$:

$$P_{xx}(z) = G_x(z)G_x^*(z^{-1}),$$  \hspace{1cm} (22.132)

where $G_x(z)$ is a causal filter, i.e., $g_x[k] = \mathcal{Z}^{-1}\{G_x(z)\} = 0, \ k < 0$.

**Example 22.4: Factorization of $P_{xx}(z)$.

Assume that both signal and noise are AR(1), i.e.,

$$S_t = \alpha S_{t-1} + AW_t, \quad N_t = \beta N_{t-1} + BW'_t,$$ \hspace{1cm} (22.133)

where $W_t$ and $W'_t$ are independent white noise (not necessarily Gaussian).

In Figure 22.9 we plot $S_t, N_t$ and $X_t = S_t + N_t$, by assuming $W_t$ and $W'_t$ are white Gaussian noise $\sim N(0, 1)$, and

$$|A|^2 = |B|^2 = 1, \quad \alpha = 2^{-1/4} \approx 0.8409, \quad \beta = 2^{-1} = 0.5.$$

Let us assume that both signal and noise are both AR(1), i.e.,
**Figure 22.9** Gauss-Markov signal $S_t$ and noise $N_t$ and their superposition
Using (22.128), the signal-to-noise-ratio (SNR) is given by

\[
\text{SNR} = \frac{\sigma_s^2}{\sigma_n^2} = \frac{|A|^2(1 - |\beta|^2)}{|B|^2(1 - |\alpha|^2)} = 2.56 \approx 4.08 \text{ dB}
\]

The power spectrums take the following form:

\[
P_{ss}(z) = \frac{|A|^2}{(1 - \alpha z^{-1})(1 - \bar{\alpha} z)}, \quad P_{nn}(z) = \frac{|B|^2}{(1 - \beta z^{-1})(1 - \bar{\beta} z)}
\]

The power spectral density

\[
P_s(\omega) = \frac{1}{2\pi} P_{ss}(e^{i\omega}) = \frac{|A|^2}{2\pi|1 - \alpha e^{-i\omega}|^2} \frac{1}{2\pi(1 - 2\Re\{\alpha\} \cos \omega + |\alpha|^2)}, \quad -\pi \leq \omega \leq \pi,
\]

and a similar expression for the noise process.
\[ P_{xx}(z) = P_{ss}(z) + P_{nn}(z) \]
\[ = \frac{|B|^2(1 - \alpha z^{-1})(1 - \overline{\alpha} z) + |A|^2(1 - \beta z^{-1})(1 - \overline{\beta} z)}{(1 - \alpha z^{-1})(1 - \overline{\alpha} z)(1 - \beta z^{-1})(1 - \overline{\beta} z)} \]
\[ = \frac{|C|^2(1 - \gamma z^{-1})(1 - \overline{\gamma} z)}{(1 - \alpha z^{-1})(1 - \overline{\alpha} z)(1 - \beta z^{-1})(1 - \overline{\beta} z)} \]

where \(|C|\) and \(\gamma\) are determined by

\[ |C|^2(1 + |\gamma|^2) = |B|^2(1 + |\alpha|^2) + |A|^2(1 + |\beta|^2) \]
\[ |B|^2 \alpha + |A|^2 \beta = |C|^2 \gamma. \]
\[ \gamma = 0.5175, \text{ and } |C|^2 = 2.3326 \]

We can factor \(P_{xx}(z)\):

\[ P_{xx}(z) = G_x(z)G_x^*(z^{-1}) \quad (22.136) \]
\[ G_x(z) = \frac{C(1 - \gamma z^{-1})}{(1 - \alpha z^{-1})(1 - \beta z^{-1})} \]

The inverse Z-transform gives

\[ g_x[k] = \frac{C(\alpha - \gamma)}{\alpha - \beta} \alpha^k + \frac{C(\gamma - \beta)}{\alpha - \beta} \beta^k, \quad k = 0, 1, 2, \ldots \]

If we choose \( C \) to be real and positive, we find \( C = 1.5273 \), and

\[ g_x[k] = 1.4489 \times 2^{-\frac{k}{4}} + 0.0784 \times 2^{-k}, \quad k = 0, 1, 2, \ldots \]

Clearly, the second term dies down four times as fast as the first term as time \( k \) progresses.

Note that all zeros \( (z = 0, \gamma) \) and poles \( (z = \alpha, \beta) \) of \( G_x(z) \) are within the unit circle \( (|z| = 1) \), so that the inverse filter \( 1/G_x(z) \) should be also physically realizable. Such a linear system is called a minimum-phase system.
Let us denote the inverse $Z$-transform of $G_x^*(z^{-1})$ as $g_x^{(-)}[k]$:

\[
g_x^{(-)}[k] = Z^{-1} \left\{ G_x^*(z^{-1}) \right\} = Z^{-1} \left\{ G_x(z^{-1}) \right\} = g_x[-k] = \overline{g}_x[-k]
\]

Therefore,

\[
g_x^{(-)}[k] = 0, \quad k > 0
\]

referred to as an anti-causal filter.

Since $P_{xx}(z) = G_x(z)G_x^*(z)$,

\[
R_{xx}[k] = g_x[k] \ast g_x^{(-)}[k] = \sum_{j=-\infty}^{0} g_x[k-j]g_x^{(-)}[j].
\]

(22.137)

(22.138)
Define $A(z)$ such that

$$P_{sx}(z) = A(z)G_x^*(z^{-1}). \quad (22.139)$$

By taking the inverse $Z$-transform,

$$R_{sx}[k] = a[k] \ast g_x^{(-)}[k] = \sum_{j=-\infty}^{0} a[k - j]g_x^{(-)}[j] \quad (22.140)$$

where

$$a[k] = Z^{-1}\{A(z)\}. \quad (22.141)$$

By substituting (22.138) and (22.140) into (22.89)

$$\sum_{k=0}^{\infty} h[k]R_{xx}[j - k] = R_{sx}[j + p], \quad \text{for all } j \geq 0. \quad (22.89)$$
\[
\sum_{j=-\infty}^{0} a[t + p - j]g_x^{(-)}[j] = \sum_{k=0}^{\infty} h[k] \sum_{j=-\infty}^{0} g_x[t - k - j]g_x^{(-)}[j], \quad t \geq 0.
\]

or
\[
\sum_{j=-\infty}^{0} g_x^{(-)}[j] \left[ a[t + p - j] - \sum_{k=0}^{\infty} h[k]g_x[t - k - j] \right] = 0, \quad t \geq 0
\]

The last equation is satisfied if
\[
a[t + p - j] = \sum_{k=0}^{\infty} h[k]g_x[t - k - j], \quad t \geq 0, \quad j < 0
\]

\[
a[t' + p] = \sum_{k=0}^{\infty} h[k]g_x[t' - k], \quad t' > 0.
\]
The $Z$-transform of LHS of (22.145)

\[
\text{LHS} = Z \{ a[t'] + p u[t'] \} = [A(z)z^p]_+ = \left[ \frac{P_{sx}(z)z^p}{G_x^*(z^{-1})} \right]_+ \triangleq F_{sx}(z). \tag{22.146}
\]

The RHS of (22.145) is the convolution of the two causal functions

\[
\text{RHS} = Z \left\{ \sum_{k=0}^{\infty} h[k]g_x[t-k] \right\} = H(z)G_x(z) \tag{22.147}
\]

From the last two equations we find

\[
H_{\text{opt}}(z) = \frac{F_{sx}(z)}{G_x(z)} = \frac{1}{G_x(z)} \left[ \frac{P_{sx}(z)z^p}{G_x^*(z^{-1})} \right]_+. \tag{22.148}
\]
**Pure prediction:** \( N_t = 0 \) for all \( t \)

\[
P_{sx}(z) = P_{ss}(z) = G_s(z)G_s^*(z^{-1})
\]

Thus, (22.148) becomes

\[
H_{opt}(z) = \frac{1}{G_s(z)} \left[ \frac{P_{ss}(z)z_p}{G_s^*(z^{-1})} \right]_+ = \frac{1}{G_s(z)} [G_s(z)z^p]_+ \tag{22.149}
\]

which is (22.117).

**Smoothing:** \( p = -d \leq 0 \).

An optimal smoothing filter with delay \( d \) is found by setting \( p = -d \) in (22.148).

**Uncorrelated signal and noise:**

\[
P_{sx}(z) = P_{ss}(z).
\]
Then, an optimum $p$-step predictor is

$$H_{opt}(z) = \frac{1}{G_x(z)} \left[ \frac{P_{ss}(z) z^p}{G_x^*(z^{-1})} \right] +$$

(22.150)

Similarly, an optimum smoothing filter is given by setting $p = -d \leq 0$.

**Example 22.5: Optimal predicting filter for uncorrelated signal and noise**

Let us consider an optimal prediction problem. In Example 22.4, we found

$$P_{xx}(z) = G_x(z) G_x^*(z^{-1})$$

where

$$G_x(z) = \frac{C(1 - \gamma z^{-1})}{(1 - \alpha z^{-1})(1 - \beta z^{-1})}$$

and

$$G_x^*(z^{-1}) = \frac{C(1 - \bar{\gamma} z)}{(1 - \bar{\alpha} z)(1 - \bar{\beta} z)}.$$
\[
\frac{P_{ss}(z) z^p}{G^*_x(z^{-1})} = \frac{|A|^2}{C} \frac{(1 - \overline{\beta}z) z^p}{(1 - \alpha z^{-1})(1 - \overline{\gamma}z)} \\
= \frac{|A|^2}{C} \frac{z^{p+1}(1 - \overline{\beta}z)}{(z - \alpha)(1 - \overline{\gamma}z)} \\
= \frac{|A|^2}{C} \left[ \frac{az^p}{1 - \alpha z^{-1}} + \frac{bz^{p+1}}{1 - \overline{\gamma}z} \right]
\]  \hspace{1cm} (22.151)

where

\[a = \frac{1 - \alpha \overline{\beta}}{1 - \alpha \overline{\gamma}}, \quad \text{and} \quad b = \frac{\overline{\gamma} - \overline{\beta}}{1 - \alpha \overline{\gamma}}\]

If \(p \geq 1\), the first term in the bracket \([\quad]\) of (22.151) contributes to the causal part, whereas the second term is all non-causal.
For $p \geq 0$, we find

$$\left[ \frac{P_{ss}(z)z^p}{G_x^*(z^{-1})} \right]_+ = \frac{a|A|^2}{C} \left[ \frac{z^p}{1 - \alpha z^{-1}} \right]_+ = \frac{a|A|^2}{C} \sum_{k=0}^{\infty} \alpha^k z^{-(k-p)} u_{k-p}$$

$$= \frac{a|A|^2\alpha^p}{C} \sum_{t=0}^{\infty} \alpha^t z^{-t} = \frac{a|A|^2\alpha^p}{C} \frac{1}{1 - \alpha z^{-1}}$$

(22.152)

Thus, from (22.150)

$$H_{opt}(z) = c \alpha^p \frac{1 - \beta z^{-1}}{1 - \gamma z^{-1}}$$

(22.153)

where

$$c = \frac{a|A|^2}{|C|^2} = \frac{(1 - \alpha \overline{\beta})|A|^2}{(1 - \alpha \overline{\gamma})|C|^2}$$
By taking the $Z$-transform, we obtain

$$h_{opt}[k] = \begin{cases} c\alpha^p, & k = 0; \\ c(\gamma - \beta)\alpha^p\gamma^{k-1}, & k = 1, 2, \ldots \end{cases}$$

which is an almost geometrically decaying function, i.e.,

$$h_{opt}[k+1]/h_{opt}[k] = \gamma < 1$$
for all $k \geq 1$, but

$$h_{opt}[1]/h_{opt}[0] = \gamma - \beta < \gamma < 1.$$

In Figure 22.11 we plot the GMP signal $S_t$ and noisy version $X_t = S_t + N_t$ and the filtered output $Y_t$, where we set $p = 0$. 
Figure 22.11 The GMP signal $S_t$, its noisy version $X_t$, and the filter output $Y_t = \hat{S}_t$ with $p = 0$. 
Let us examine the above solution for two special cases.

**White noise:** \( \beta = 0 \), then (22.135) becomes \( P_{nn}(z) = B \)

The predicting filter is found from (22.153)

\[
H_{\text{opt}}(z) = \frac{c\alpha^p}{1 - \gamma z^{-1}}
\]

Thus,

\[
h_{\text{opt}}[k] = \begin{cases} 
  c\alpha^p\gamma^k, & k = 0, 1, 2, \ldots \\
  0, & k < 0,
\end{cases}
\]

(22.155)

**Pure prediction:** By setting \( B = 0 \), we have

\[
P_{xx}(z) = P_{ss}(z) = \frac{|A|^2}{(1 - \alpha z^{-1})(1 - \overline{\alpha} z)}
\]

\[
G_s(z) = \frac{A}{1 - \alpha z^{-1}}, \quad G_s^*(z^{-1}) = \frac{\overline{A}}{1 - \overline{\alpha} z}
\]
Thus, from (22.149)

\[ H_{\text{opt}}(z) = (1 - \alpha z^{-1}) \left[ \frac{z^p}{1 - \alpha z^{-1}} \right] + \]

Following the derivation step in (22.152),

\[ \left[ \frac{z^p}{1 - \alpha z^{-1}} \right]_+ = \frac{\alpha^p}{1 - \alpha z^{-1}} \]

Thus,

\[ H_{\text{opt}}(z) = (1 - \alpha z^{-1}) \frac{\alpha^p}{1 - \alpha z^{-1}} = \alpha^p \]

an “attenuator” by the factor \( \alpha^p \), as obtained in (22.125)
Evaluation of minimum mean square errors:

\[
\mathcal{E}_{\text{min}} = E \left[ \left| S_{t+p} - \hat{S}_{t+p} \right|^2 \right]
\]  
(22.156)

By substituting

\[
\hat{S}_t = \sum_{k=0}^{\infty} h_{\text{opt}}[k] X_{t-k}
\]  
(22.157)

\[
\mathcal{E}_{\text{min}} = R_{ss}[0] - 2\Re \left\{ \sum_{j=0}^{\infty} R_{sx}[j+p] h^*_{\text{opt}}[j] \right\}
\]

\[
+ \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} h_{\text{opt}}[k] R_{xx}[j-k] h^*_{\text{opt}}[j]
\]  
(22.158)

By substituting

\[
R_{sx}[j+p] = \sum_{k=0}^{\infty} h_{\text{opt}}[k] R_{xx}[j-k]
\]  
(22.159)
into (22.158) we have

\[ \mathcal{E}_{\text{min}} = R_{ss}[0] - \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} h_{\text{opt}}[k] R_{xx}[j - k] h^{*}_{\text{opt}}[j] \]  

(22.160)

Using Parseval’s formula, we can write the above (Problem 22.21)

\[ \mathcal{E}_{\text{min}} = \frac{1}{2\pi i} \oint P_{ss}(z) \frac{dz}{z} - \frac{1}{2\pi i} \oint H_{\text{opt}}(z) P_{xx}(z) H_{\text{opt}}^{*}(z^{-1}) \frac{dz}{z} \]  

(22.161)

By substituting (22.148) into the above and using \( G_x(z) G_x^{*}(z^{-1}) = P_{xx}(z) \), we finally obtain

\[ \mathcal{E}_{\text{min}} = \frac{1}{2\pi i} \oint P_{ss}(z) \frac{dz}{z} - \frac{1}{2\pi i} \oint F_{sx}(z) \bar{F}_{sx}(z^{-1}) \frac{dz}{z} , \]

(22.162)
where $F_{sx}(z)$ was defined in (22.146):

$$F_{sx}(z) = \left[ \frac{P_{sx}(z)z^p}{G^*_x(z^{-1})} \right]_+.$$ (22.163)

**Pure prediction case:**

$$P_{sx}(z) = P_{xx}(z) = P_{ss}(z) = G_s(z)G^*_s(z^{-1})$$

Then

$$g_s[t] = 0, \quad t < 0$$

Then

$$\mathcal{E}_{\text{min}} = \int G_s(z)G^*_s(z^{-1}) \frac{dz}{z} - \int F_s(z) \overline{F_s(z^{-1})} \frac{dz}{z}$$ (22.164)

$F_s(z)$ is optimum predictor for the whitened signal as in Figure 22.6:

$$F_s(z) = \left[ G_s(z)z^p \right]_+$$
Using the inverse formula \( \mathcal{Z}^{-1} \{ G_s(f) z^p \} = g_s[k + p] \),

\[
f_s[k] = \mathcal{Z}^{-1} \{ F_s(z) \} = g_s[k + p] u[k]
\]

Applying the Parseval’s formula again,

\[
E_{\text{min}} = \sum_{k=0}^{\infty} |g_s[k]|^2 - \sum_{k=0}^{\infty} |f_s[k]|^2 = \sum_{k=0}^{p-1} |g_s[k]|^2. \tag{22.165}
\]

In referring to Figure 22.6 (a), the MMSE of the pure prediction is given by the norm square of the noncausal part of the second impulse response \( g_s[g + p] \).
Figure 22.6 (a) The relation between $g_s[k]$ and $f_s[k]$; (b) An optimal predictor $f_s[k] = g_s[k + p]$ that acts on white noise.
Estimation Theory: Preliminary

18.1 Parameter Estimation

We consider random variable (RV) $X$ with probability distributions $F(x; \theta)$ with parameter $\theta$,

$$\theta = (\theta_1, \theta_2, \ldots, \theta_M).$$

The value of parameter $\theta$ is unknown and we want to estimate it from observations

$$x = (x_1, x_2, \ldots, x_n) \in \mathcal{X}$$

drawn from the distribution $F(x; \theta)$. We want to find a function $T(\cdot)$ such that

$$\hat{\theta} = T(x)$$

is as close to $\theta$ as possible.
Definition 18.1 (Unbiasedness, efficiency and consistency of an estimator).

1. An estimator $\hat{\theta}(X)$ is said to be unbiased, if $E[\hat{\theta}(X)] = \theta$; otherwise, it is called biased. The bias is defined as

$$b(\theta) = E[\hat{\theta}(X)] - \theta.$$  \hfill (18.1)

2. An unbiased estimator $\hat{\theta}^*(X)$ is said to be efficient, if it is a minimum-variance estimator, that is, $\text{Var}[\hat{\theta}^*(X)] \leq \text{Var}[\hat{\theta}(X)]$ for any other unbiased estimator $\hat{\theta}(X)$.

3. A sequence of estimators is said to be consistent, if the sequence converges in probability to $\theta$. \hfill $\square$

Definition 18.2 (Sufficient statistic). A statistic $T(X)$ is said to be sufficient for parameter $\theta$, if the conditional probability density (or mass) function of $X$, given $T(X) = t$, does not depend on $\theta$. \hfill $\square$

This means that, given $T(x) = t$, full knowledge of the measurement $x$ does not bring any additional information concerning $\theta$. 

12/04/2013

Copyright Hisashi Kobayashi 2013
18.1.2 Maximum-Likelihood Estimation

\[ L_x(\theta) \triangleq \begin{cases} f_X(x; \theta), & \text{for continous RV } X, \\ p_X(x; \theta), & \text{for discrete RV } X. \end{cases} \quad (18.24) \]

is called the **likelihood function**: Any value of \( \theta \) that maximizes the likelihood function is called a **maximum-likelihood estimate** (MLE),

\[ \hat{\theta} = \arg \max_{\theta} L_x(\theta). \quad (18.25) \]

The procedure to find an MLE is called **maximum-likelihood estimation**. If the likelihood function is differentiable with respect to its parameter, a necessary condition for an MLE to satisfy is

\[ \nabla_{\theta} L_x(\theta) = 0, \quad \text{i.e.,} \quad \frac{\partial L_x(\theta)}{\partial \theta_m} = 0, \quad m = 1, 2, \ldots, M. \quad (18.26) \]