7 Sensitivity Analysis

A recurrent theme underlying methodology for analysis in the presence of missing data is the need to make assumptions that cannot be verified based on the observed data. If the assumption of missing at random is adopted, as in the methods in Chapters 3–5, the untestable assumption is that the probability of observing any missing data pattern $R = r$, conditional on the full data $Z$, depends only on components of $Z$ that are observed under this pattern. Alternatively, under a pattern mixture model formulation, the analyst must make some sort of untestable identifiability assumption, which induces an assumption about the nature of missing data.

In any statistical modeling and analysis context, it is natural to be concerned about robustness or sensitivity of inferences to departures from assumptions. Thus, there has been considerable recent interest in the case of modeling and analysis methods in the presence of missing data in what has been referred to as sensitivity analysis. A sensitivity analysis focuses on examining and perhaps quantifying somehow the effect of departures from assumptions. The extent to which conclusions are stable (or not) across a range of departures can provide the analyst with important information on how much faith to attach to the results.

In any of the approaches we have discussed, in addition to unverifiable assumptions about the missing data mechanism or assumption made to achieve identifiability, assumptions are also made on other aspects, such as the adoption of models for the full data. In this chapter, we focus on specifically on sensitivity to assumptions on the missing data mechanism.

While approaches for modeling and analysis under untestable assumptions, in particular MAR, are well developed, the literature on sensitivity analysis is still evolving, and no consensus view has emerged on how best to carry out such analyses. Accordingly, our discussion here is limited to illustrating in the context of a few simple settings principles that underlie formal approaches to sensitivity analysis. This demonstration will thus provide a foundation for further study of this evolving literature.

Our presentation is based on that in Chapter 5 of the National Research Council (2010) report The Prevention and Treatment of Missing Data in Clinical Trials, which contains more details.
Departures from a MAR assumption naturally involve a **missing not at random** (MNAR) mechanism under which the probability of any missingness pattern given the full data \( Z \) depends on components of \( Z \) that are not observed under that pattern. Clearly, different forms of such dependence, which are not observable, could represent the observed data equally well. This suggests inspecting stability of results across a range of plausible forms of dependence. This is the theme behind the principles discussed in this chapter.

### 7.1 Estimation of a single mean

We consider first the simplest possible setting discussed in **EXAMPLE 1** of Section 1.4, that of estimation of the mean of a single random variable.

Here, the full data are \( Z = Y \), where \( Y \) is a scalar outcome, and, as usual, we define \( R = 1 \) if \( Y \) is observed and \( R = 0 \) otherwise. Then the observed data can be written as \((R, RY)\), and the observed data from a sample of \( N \) individuals are \((R_i, R_iY_i), i = 1, \ldots, N\). The goal is to estimate

\[
\mu = E(Y)
\]

based on the observed sample data.

As we discussed in **EXAMPLE 1** of Section 1.4, if the missingness mechanism is **MCAR**, \( pr(R = 1 | Y) \) does not depend on \( Y \), so that

\[
pr(R = 1 | Y) = pr(R = 1) = \pi,
\]
say, and \( pr(R = 0 | Y) = pr(R = 0) = 1 - \pi \), so that \( R \perp Y \). Then the **complete case** estimator (1.18),

\[
\hat{\mu}_c = \frac{\sum_{i=1}^{N} R_i Y_i}{\sum_{i=1}^{N} R_i},
\]

is a **consistent estimator** for \( \mu \).

However, if instead

\[
pr(R = 1 | Y) = \pi(Y),
\]

say, so that \( pr(R = 0 | Y) = 1 - \pi(Y) \) depends on \( Y \), which is unobserved when \( R = 0 \), then the missingness mechanism is **MNAR**. We demonstrated in **EXAMPLE 1** of Section 1.4 that, under MNAR, \( \hat{\mu}_c \) is an **inconsistent** estimator for \( \mu \) in general.
Here, the only two possible missingness mechanisms are MCAR (which is the simplest form of MAR) and MNAR, where the assumption of MCAR is unverifiable from the observed data. A sensitivity analysis would thus examine the effect of deviations from MCAR on inference on $\mu$.

**ILLUSTRATION OF AN UNTESTABLE ASSUMPTION:** In this simple setting, the nature of an unverifiable assumption is readily exhibited. Note that $\mu$ can be written as the **weighted average**

$$
\mu = E(Y|R = 1)\ pr(R = 1) + E(Y|R = 0) \{1 - pr(R = 1)\}.
$$

(7.1)

In (7.1), **without** any assumptions, $E(Y|R = 1)$ and $pr(R = 1)$ can be estimated from the observed data; in particular, $E(Y|R = 1)$ can be estimated by $\hat{\mu}^c$ and $pr(R = 1)$ by the **sample proportion** $\sum_{i=1}^N R_i/N$.

However, $E(Y|R = 0)$ **cannot** be estimated from the observed data. To appreciate the implication for learning about $\mu$ from the observed data, suppose that $Y$ has bounded support $[y_{\text{min}}, y_{\text{max}}]$. Then all we can say is that

$$
y_{\text{min}} \leq E(Y|R = 0) \leq y_{\text{max}},
$$

in which case, from (7.1), the value of $\mu$ can range from

$$
E(Y|R = 1) pr(R = 1) + y_{\text{min}} \{1 - pr(R = 1)\} \quad \text{to} \quad E(Y|R = 1) pr(R = 1) + y_{\text{max}} \{1 - pr(R = 1)\},
$$

(7.2)

and there is no information in the data that allows us to distinguish the value of $\mu$ within this range. Indeed, if $Y$ has unbounded support, then the situation is even more dire, and there are no bounds that can be placed on inference on $\mu$ from the observed data.

This demonstrates that $\mu$ cannot be **identified** from the observed data without making **some sort of assumption** on $E(Y|R = 0)$. Under the MCAR (MAR) assumption $R \perp \perp Y$,

$$
E(Y|R = 1) = E(Y|R = 0) = \mu,
$$

(7.3)

in which case $\mu$ can be identified, and the complete case estimator $\hat{\mu}^c$, which estimates $E(Y|R = 1)$, is then a consistent estimator for $\mu$. However, this or any other assumption on $E(Y|R = 0)$ is **untestable** based on the observed data.
We now consider approaches to sensitivity analysis that are suggested by the foregoing developments.

**PATTERN MIXTURE MODEL APPROACH:** Note that (7.1) is the simplest version of a pattern mixture model. Following the same principles discussed in Chapter 6, we can make an assumption linking \( E(Y|R = 1) \) and \( E(Y|R = 0) \) and thereby identify \( E(Y|R = 0) \). In doing so, a framework for sensitivity analysis is introduced, as we now demonstrate.

To link \( E(Y|R = 1) \) and \( E(Y|R = 0) \), assume that

\[
E(Y|R = 0) = E(Y|R = 1) + \Delta. \tag{7.4}
\]

Under (7.4), for any fixed value of \( \Delta \), it is clearly possible to estimate \( E(Y|R = 0) \) from the observed data by adding \( \Delta \) to the estimate of \( E(Y|R = 1) \).

A generalization of (7.4) that accommodates different types of outcomes \( Y \) is

\[
E(Y|R = 0) = g^{-1}[g\{E(Y|R = 1)\} + \Delta], \tag{7.5}
\]

where \( g(\cdot) \) is a strictly increasing function mapping values from the range of \( Y \) to the real line. For example, \( g(u) = u \) as in (7.4) is natural for a continuous outcome, while \( g(u) = \logit(u) = \log\{u/(1-u)\} \) is appropriate for binary outcome, in which case \( \Delta \) is the log odds ratio comparing the odds of \( Y = 1 \) among individuals with \( R = 0 \) and \( R = 1 \).

In (7.4) and more generally (7.5), \( \Delta \) can be viewed as a sensitivity parameter. Each value of \( \Delta \) corresponds to a different (unverifiable) assumption about \( E(Y|R = 0) \). Substituting in (7.1), we have

\[
\mu = E(Y|R = 1) \cdot \text{pr}(R = 1) + g^{-1}[g\{E(Y|R = 1)\} + \Delta] \cdot \{1 - \text{pr}(R = 1)\}. \tag{7.6}
\]

Note that \( \mu = \mu(\Delta) \) in (7.6) can be viewed as a function of \( \Delta \), and, for any fixed \( \Delta \), \( \mu \) can then be estimated by substituting the estimates of \( E(Y|R = 1) \) and \( \text{pr}(R = 1) \) based on the data.

In particular, recalling that \( \hat{\mu}^c \) is an estimator for \( E(Y|R = 1) \) and \( \hat{\pi} = \sum_{i=1}^N R_i/N \) is an estimator for \( \text{pr}(R = 1) \), in the case where \( g(u) = u \), for example, for fixed \( \Delta \), the estimator \( \hat{\mu}(\Delta) \), say, is

\[
\hat{\mu}(\Delta) = \hat{\mu}^c \hat{\pi} + (\hat{\mu}^c + \Delta)(1 - \hat{\pi}) = \hat{\mu}^c + \Delta(1 - \hat{\pi}). \tag{7.7}
\]

It is straightforward to derive approximate large sample standard errors for \( \hat{\mu}(\Delta) \) and associated confidence intervals for \( \mu(\Delta) \).
Taking $\Delta = 0$ in (7.6) yields (7.3) and thus

$$
\mu = E(Y|R = 1) \Pr(R = 1) + E(Y|R = 1) \{1 - \Pr(R = 1)\} = E(Y|R = 1),
$$

which corresponds to MCAR/MAR from above. This suggests that a sensitivity analysis to the MAR assumption by obtaining estimates $\tilde{\mu}(\Delta)$ and corresponding confidence intervals over a range of $\Delta$ values that includes $\Delta = 0$. The extent to which the estimator and its confidence interval change across a plausible range of value of $\Delta$ reflects the sensitivity of inferences to departures from the MCAR/MAR assumption.

**SELECTION MODEL APPROACH:** Another approach is to introduce instead a sensitivity parameter in a model for the missingness mechanism. For example, consider the model

$$
\logit\{\Pr(R = 1|Y)\} = \alpha + \delta Y. \quad (7.8)
$$

In (7.8), $\delta = 0$ corresponds to MCAR, while when $\delta \neq 0$ the odds ratio associated with a unit change in $Y$ is

$$
\frac{\Pr(R = 1|Y = y + 1)/\Pr(R = 0|Y = y + 1)}{\Pr(R = 1|Y = y)/\Pr(R = 0|Y = y)} = \exp(\delta).
$$

Thus, if $\delta > 0$, the probability that $Y$ is not missing increases with the value of $Y$, and vice versa. The magnitude of $\delta$ then dictates the extent of departure from MCAR, making $\delta$ a natural sensitivity parameter.

Of course, we cannot estimate $\alpha$ and $\delta$ in (7.8) jointly from the observed data because $Y$ is missing whenever $R = 0$. However, if we could fit (7.8) to the observed data for fixed $\delta$ and then use the fitted (7.8) to estimate $\mu$, e.g., via inverse probability weighting, we would have the basis for a sensitivity analysis to departures from MCAR in which we vary $\delta$ across a plausible range including $\delta = 0$.

Accordingly, we first examine if, assuming (7.8), we can estimate $\alpha$ and consequently $\mu$ for fixed $\delta$. Write (7.8) equivalently as

$$
\Pr(R = 1|Y) = \expit(\alpha + \delta Y) = \pi(Y; \alpha, \delta),
$$

say, where as usual $\expit(u) = e^u/(1 + e^u)$. 
Assuming that the model (7.8) is correctly specified, it is of course straightforward that
\[
E_{\alpha, \delta} \left\{ \frac{R}{\pi(Y; \alpha, \delta)} \right\} = 1. \tag{7.9}
\]

If we write
\[
E \left\{ \frac{R}{\pi(Y; \alpha, \delta)} \right\} = q(\alpha, \delta),
\]
then, for any fixed \( \delta \), then, under (7.8), the true value \( \alpha_0 \), say, of \( \alpha \) is such that
\[
q(\alpha_0, \delta) = 1.
\]

In this case, for fixed \( \delta \), we have
\[
\mu = E \left\{ \frac{RY}{\pi(Y; \alpha_0, \delta)} \right\}, \tag{7.10}
\]
which can be estimated from the observed data.

Consequently, if we are willing to entertain the model in (7.8), for fixed \( \delta \), consistent estimators for \( \alpha \) and thereby \( \mu \) can be obtained from the empirical counterparts of (7.9) and (7.10) based on the observed data \((R_i, R_i Y_i), i = 1, \ldots, N\). Namely, noting that
\[
1/\pi(Y; \alpha, \delta) = 1 + \exp\{- (\alpha + \delta Y_i)\}, \tag{7.11}
\]
an estimator \( \hat{\alpha}(\delta) \) for \( \alpha \) for fixed \( \delta \) can be obtained by solving in \( \alpha \)
\[
\sum_{i=1}^{N} \left( R_i [1 + \exp\{- (\alpha + \delta Y_i)\}] - 1 \right) = 0, \tag{7.12}
\]
where we emphasize dependence on the fixed \( \delta \), from whence it follows that an estimator for \( \mu \) for fixed \( \delta \) is
\[
\hat{\mu}(\delta) = N^{-1} \sum_{i=1}^{N} R_i Y_i \left( 1 + \exp[-\{\hat{\alpha}(\delta) + \delta Y_i\}] \right). \tag{7.13}
\]

By algebra, from (7.12)
\[
\exp\{- \hat{\alpha}(\delta)\} = \frac{\sum_{i=1}^{N} (1 - R_i)}{\sum_{i=1}^{N} R_i \exp(-\delta Y_i)},
\]
so that, from (7.13)
\[
\hat{\mu}(\delta) = N^{-1} \sum_{i=1}^{N} R_i Y_i \left\{ 1 + \frac{\sum_{\ell=1}^{N} (1 - R_\ell)}{\sum_{\ell=1}^{N} R_\ell \exp(-\delta Y_\ell)} \right\} \exp(-\delta Y_i).
\]
In fact, note that, writing (7.12) and (7.13) for fixed $\delta$ as the set of Raiders of missing data

$$\sum_{i=1}^{N} \{ R_i Y_i \left( 1 + \exp \{ - \{ \alpha(\delta) + \delta Y_i \} \} \} - \mu(\delta) \} = 0$$

$$\sum_{i=1}^{N} \sum_{i=1}^{N} \{ R_i \left( 1 + \exp \{ - \{ \alpha(\delta) + \delta Y_i \} \} \right) - 1 \} = 0,$$

it is clear that $\hat{\alpha}(\delta)$ and $\hat{\mu}(\delta)$ are $M$-estimators and hence are asymptotically normal with asymptotic covariance matrix that can be estimated by the usual sandwich estimator deduced from the general formulation in (1.42).

The corresponding sensitivity analysis can be based on the foregoing results. This would entail obtaining $\hat{\mu}(\delta)$ and a corresponding confidence interval over a plausible range of values for $\delta$ about the value $\delta = 0$ (MCAR/MAR). The extent to which the estimator and its confidence interval change as $\delta$ changes provide an indication of the sensitivity of inferences to departures from the MCAR assumption.

**IMPLICATION:** It is of interest to deduce the range of $\mu(\delta)$ values induced from the distribution of the observed data as $\delta \to \pm \infty$.

Assume that the support of $Y$ is $[y_{\text{min}}, y_{\text{max}}]$ as before. Then the distribution of the observed data can be characterized by

$$\pi = \text{pr}(R = 1) \quad \text{and} \quad p_{Y|Y}(y|1),$$

where $p_{Y|Y}(y|r)$ is the conditional density of $Y$ given $R$ evaluated at $R = r$, and $y_{\text{min}} \leq y \leq y_{\text{max}}$.

Consequently,

$$q(\alpha, \delta) = E \left\{ \frac{R}{\pi(Y; \alpha, \delta)} \right\} = E \left[ E \left\{ \frac{R}{\pi(Y; \alpha, \delta)} \ \mid \ R \right\} \right] = \left\{ \int_{y_{\text{min}}}^{y_{\text{max}}} \frac{1}{\pi(Y; \alpha, \delta)} p_{Y|Y}(y|1) \ dy \right\} \pi$$

$$= \left( \int_{y_{\text{min}}}^{y_{\text{max}}} \left[ \exp \{ - (\alpha + \delta) y \} + 1 \right] p_{Y|Y}(y|1) \ dy \right) \pi$$

$$= \left[ \exp(-\alpha) \int_{y_{\text{min}}}^{y_{\text{max}}} \exp(-\delta y) p_{Y|Y}(y|1) \ dy + 1 \right] \pi$$

Setting $q(\alpha, \delta) = 1$, we then obtain that $\alpha(\delta)$ satisfies

$$\exp \{ - \alpha(\delta) \} \int_{y_{\text{min}}}^{y_{\text{max}}} \exp(-\delta y) p_{Y|Y}(y|1) \ dy = \left( \frac{1}{\pi} - 1 \right). \quad (7.14)$$
Similarly,
\[
\mu(\delta) = E\left\{ \frac{RY}{\pi(Y; \alpha, \delta)} \right\} = \left( \int_{y_{\min}}^{y_{\max}} y \left[ \exp\{-(\alpha + \delta y)\} + 1 \right] p_{Y|R}(y|1) \, dy \right) \pi
\]
\[= \left[ \exp(-\alpha) \int_{y_{\min}}^{y_{\max}} y \exp(-\delta y) p_{Y|R}(y|1) \, dy + E(Y|R = 1) \right] \pi,
\]
which we can write as
\[
E(Y|R = 1) \pi + \left[ \exp(-\alpha) \int_{y_{\min}}^{y_{\max}} y \exp(-\delta y) p_{Y|R}(y|1) \, dy \right] \pi.
\]
Using (7.14), it follows that (7.15) is equal to
\[
E(Y|R = 1) \pi + \left( \frac{1}{\pi} - 1 \right) \pi \left\{ \int_{y_{\min}}^{y_{\max}} y \exp(-\delta y) p_{Y|R}(y|1) \, dy \right\}. \tag{7.16}
\]
The term in braces in (7.16) is a **weighted average** of \( Y \) from \( y_{\min} \) to \( y_{\max} \), where the weighting is by \( \exp(-\delta y) p_{Y|R}(y|1) \). Thus, as \( \delta \to \infty \), the mass of the weighting **converges** near the minimum value \( y_{\min} \), so that the term in braces converges to \( y_{\min} \). It follows that
\[
\mu(\delta) \to E(Y|R = 1) \pi + (1 - \pi) y_{\min}.
\]
Similarly, as \( \delta \to -\infty \),
\[
\mu(\delta) \to E(Y|R = 1) \pi + (1 - \pi) y_{\max}.
\]
These are exactly the same limits deduced earlier directly from the **pattern mixture perspective** in (7.2), which were sharp bounds.

### 7.2 Estimation of a single mean with auxiliary data

Consider now the case where we continue to be interested in the mean \( \mu = E(Y) \) for a scalar outcome \( Y \) that can be missing, but additional **auxiliary data** \( V \) are always available. That is, the full data are now \( Z = (Y, V) \), and \( R = (R_1, R_2)^T \), where \( R \) takes on the two possible values \( R = (1, 1)^T \) when \( Y \) is observed and \( R = (0, 1)^T \) when \( Y \) is missing. As in **EXAMPLE 1** of Section 1.4 and Section 5.1, let \( C = 1 \) if \( R = (1, 1)^T \), indicating a **complete case**, and \( C = 0 \) if \( R = (0, 1)^T \), indicating that \( Y \) is **missing**.
The observed data are then \((C, CY, V)\), and the observed sample data are \((C_i, C_iY_i, V_i), i = 1, \ldots, N\). From Sections 1.4 and 5.1, we have seen that it is possible to obtain a consistent estimator for \(\mu\) under the assumption that, as in (1.21) and (5.2), missingness of \(Y\) depends only on \(V\) and not on the possibly unobserved \(Y\),

\[
\operatorname{pr}(C = 1|Y, V) = \operatorname{pr}(C = 1|V) = \pi(V),
\]

so that \(C \perp Y|V\), and the missingness mechanism in MAR. Equivalently, (7.17) implies that

\[
E(Y|V, C = 1) = E(Y|V, C = 0).
\]

Of course, the MAR assumption is unverifiable based on the observed data, and it is not possible to estimate (identify) \(E(Y|V, C = 0)\) in (7.18) based on the observed data without such an assumption.

We now sketch how sensitivity analyses to the MAR assumption can be obtained.

**PATTERN MIXTURE MODEL APPROACH:** As in the situation with no auxiliary data in the preceding section, we can make an assumption linking \(E(Y|V, C = 1)\) and \(E(Y|V, C = 0)\) and identify \(E(Y|V, C = 0)\). With \(g(\cdot)\) a strictly increasing function as before, analogous to (7.5), one can specify

\[
E(Y|V, C = 0) = g^{-1}[g\{\nu(V)\} + \Delta],
\]

say, where \(\nu(V) = E(Y|V, C = 1)\). As in (7.4) and (7.5), \(\Delta\) is a sensitivity parameter, and each value of \(\Delta\) represents a different, untestable assumption about \(E(Y|V, C = 0)\), with \(\Delta = 0\) corresponding to MAR as in (7.18).

To illustrate how the sensitivity analysis would proceed, suppose that \(Y\) is continuous and \(g(u) = u\). For definiteness, one must specify the relationship between \(Y\) and \(V\) for \(C = 1\) represented by \(\nu(V)\). For example, a simple choice of \(\nu(\cdot)\) is a linear regression model; i.e.,

\[
\nu(V; \gamma) = E(Y|V, C = 1) = \gamma_0 + \gamma_1^T V.
\]

The regression model (7.20) can be fitted based on the observed sample data for individuals for whom \(C = 1\); that is, with both \(Y\) and \(V\) observed, yielding estimators \(\hat{\gamma}_0\) and \(\hat{\gamma}_1\).

With this choice of \(g(\cdot)\) and the specification (7.20), the linkage relationship (7.19) becomes

\[
E(Y|V, C = 0) = \gamma_0 + \gamma_1^T V + \Delta.
\]

164
Under (7.21), for any $V$ and fixed $\Delta$, an estimator for $E(Y|V, C = 0)$ can be derived as

$$\hat{\gamma}_0 + \hat{\gamma}_1^T V + \Delta,$$

so that $E(Y|V, C = 0)$ is **imputed** as the predicted value of $E(Y|V, C = 1)$ plus the shift $\Delta$.

**REMARKS:**

- The linkage formulation (7.19) assumes that the difference between $E(Y|V, C = 0)$ and $E(Y|V, C = 1)$ depends on a **constant** $\Delta$ for all $V$ and thus does not depend on $V$. A more flexible specification would allow such dependence; for example, for general $g(\cdot)$,

$$E(Y|V, C = 0) = g^{-1}[g(\nu(V)) + (\Delta_0 + \Delta_1^T V)].$$

For continuous $Y$ and $g(u) = u$, this becomes

$$E(Y|V, C = 0) = \nu(V) + (\Delta_0 + \Delta_1^T V).$$

In either case, $(\Delta_0, \Delta_1^T)^T$ is a bivariate **sensitivity parameter** characterizing the **sensitivity function** $(\Delta_0 + \Delta_1^T V)$.

In principle, any general **sensitivity function** $s(V; \Delta)$ could be specified. It is important to recognize that the form of $s(V; \Delta)$ **cannot be validated** based on the observed data.

- The choice of model for $\nu(V) = E(Y|V, C = 1)$ can also be made more **flexible** than the simple linear specification in (7.20). Note that the form of $\nu(V)$ **can be validated** based on the observed data, as it characterizes the relationship between $Y$ and $V$ among individuals for whom both are observed ($C = 1$).

A similar formulation in the case where $Y$ is **binary** would proceed in the obvious manner with $g(u) = \logit(u)$, where a model such as

$$\logit(\nu(V; \gamma)) = \gamma_0 + \gamma_1^T V$$

for $E(Y|V, C = 1)$ would be reasonable.
**IMPLEMENTATION:** The sensitivity analysis based on the foregoing developments would then proceed as follows. Let \( \pi = \text{pr}(C = 1) \), which of course can be estimated by \( \hat{\pi} = N^{-1} \sum_{i=1}^{N} C_i \).

1. Specify a model \( \nu(V) = \nu(V; \gamma) \) for \( E(Y|V, C = 1) \) and a sensitivity function \( s(V; \Delta) \). Fit \( \nu(V; \gamma) \) to the observed data from individuals for whom \( Y \) is observed \( (C = 1) \) to obtain \( \hat{\gamma} \).

2. Analogous to (7.6), using the fact that

\[
E(Y|C = c) = E\{E(Y|V, C = c)|C = c\}
\]

for \( c = 0, 1 \) and assuming that the model \( \nu(V) \) is correctly specified, we have

\[
\mu = \pi E(Y|C = 1) + (1 - \pi) E(Y|C = 0)
= \pi E\{\nu(V)|C = 1\} + (1 - \pi) E\left(g^{-1}[g\{\nu(V)\} + s(V; \Delta)]|C = 0\right).
\]

(7.22)

To obtain an estimator for \( \mu \) based on (7.22), one must substitute estimators for each component on the right hand side as follows.

Given the fitted model \( \nu(V; \hat{\gamma}) \), \( E\{\nu(V)|C = 1\} \) in the first term on the right hand side of (7.22) can be estimated by the sample mean

\[
\frac{\sum_{i=1}^{N} C_i \nu(V_i; \hat{\gamma})}{\sum_{i=1}^{N} C_i}.
\]

(7.23)

The second term on the right hand side of (7.22) likewise can be estimated by the sample mean

\[
\frac{\sum_{i=1}^{N} (1 - C_i) g^{-1}[g\{\nu(V_i; \hat{\gamma})\} + s(V_i; \Delta)]}{\sum_{i=1}^{N} (1 - C_i)};
\]

(7.24)

note that in (7.24) \( E(Y|V, C = 1) \) is being imputed for individuals \( i \) for whom \( C_i = 0 \).

The estimator for \( \mu \) for fixed \( \Delta \), \( \hat{\mu}(\Delta) \), say, is then obtained by substituting \( \hat{\pi} \), (7.23), and (7.24) into (7.22). Evidently, it is possible to derive asymptotic standard errors for \( \hat{\mu}(\Delta) \) and associated confidence intervals.
In the case of \( g(u) = u \), for example, the estimator is

\[
\hat{\mu}(\Delta) = \frac{\sum_{i=1}^{N} C_i \nu(V_i; \hat{\gamma})}{\sum_{i=1}^{N} C_i} + (1 - \hat{\pi}) \frac{\sum_{i=1}^{N} (1 - C_i) \{\nu(V_i; \hat{\gamma}) + s(V_i; \Delta)\}}{\sum_{i=1}^{N} (1 - C_i)},
\]

(7.25)

where we emphasize that (7.25) depends on the fixed \( \Delta \).

When \( \Delta = 0 \), so that \( s(V; \Delta) = 0 \), corresponding to MAR, (7.25) becomes

\[
\hat{\mu}(0) = N^{-1} \sum_{i=1}^{N} \nu(V_i; \hat{\gamma}),
\]

the sample average of the fitted model \( \nu(V; \gamma) \) for \( E(Y|V, C = 1) \). Under MAR, \( E(Y|V, C = 1) = E(Y|V, C = 0) \), and this is a sensible estimator for \( E(Y) \) under the assumption that the model is correctly specified.

As in the simpler case in Section 7.1, it follows that a sensitivity analysis to the MAR assumption would involve calculating estimates \( \hat{\mu}(\Delta) \) and corresponding confidence intervals over a range of \( \Delta \) values that includes \( \Delta = 0 \). The extent to which the estimator and its confidence interval change across a plausible range of value of \( \Delta \) reflects the sensitivity of inferences to departures from the MAR assumption.

**SELECTION MODEL APPROACH:** Analogous to the development in Section 7.1, we can instead incorporate a sensitivity parameter(s) in a model for the missingness mechanism.

Recall from (7.17) that the MAR assumption is

\[
\text{pr}(C = 1|Y, V) = \text{pr}(C = 1|V) = \pi(V),
\]

so that \( C \perp Y|V \), and it is of course possible to estimate \( \text{pr}(C = 1|V) \) from the observed data. The MAR assumption can be written equivalently as

\[
\logit\{\text{pr}(C = 1|Y, V)\} = \logit\{\text{pr}(C = 1|V)\} = \alpha(V),
\]

(7.26)

say, so that

\[
\pi(V) = \frac{\exp\{\alpha(V)\}}{1 + \exp\{\alpha(V)\}}.
\]

Here, \( \pi(V) \) and equivalently \( \alpha(V) \) characterize the nature of the MAR assumption.

Generalizing (7.8), we can consider a model of the form

\[
\logit\{\text{pr}(C = 1|Y, V)\} = \alpha(V) + \delta(Y, V)
\]

(7.27)

for some fixed sensitivity function \( \delta(Y, V) \).
For example, a simple such sensitivity function depends only on the potentially unobserved $Y$,
\[ \delta(Y, V) = \delta Y \] (7.28)
for fixed value of a \textit{sensitivity parameter} $\delta$. Of course, more exotic sensitivity functions involving a vector of sensitivity parameters $\delta$ and both $Y$ and $V$ could be selected.

Likewise, one would typically adopt a \textit{parametric model} for $\alpha(V)$; for example, one might adopt a \textit{linear model}, e.g.,
\[ \alpha(V) = \alpha_0 + \alpha_1 V; \] (7.29)
more general nonlinear parametric models are of course possible.

As in the simpler case in the previous section, for \textit{fixed} $\delta(Y, V)$ in (7.27), $\alpha(V)$ and $\mu = E(Y)$ are \textit{identifiable} from the observed data. To see this, define from the model in (7.27),
\[ \pi(Y, V; \alpha, \delta) = \expit\{\alpha(V) + \delta(Y, V)\}, \] (7.30)
from which, analogous to (7.11),
\[ \frac{1}{\pi(Y, V; \alpha, \delta)} = 1 + \exp[-\{\alpha(V) + \delta(Y, V)\}]. \]

Then, analogous to (7.9) in the simpler setting in Section 7.1, assuming that the model selected for (7.27) is \textit{correctly specified}, we have
\[ \frac{C}{\pi(Y, V; \alpha, \delta)} \bigg| V = v \bigg] = 1 \] (7.31)
(verify). Writing (7.31) as
\[ E \left[ E \left\{ \frac{C}{\pi(Y, V; \alpha, \delta)} \bigg| R, V = v \bigg\} \bigg| V = v \right\} = 1 \text{ for all } v, \]
and letting $p_{Y|V,C}(y|v, c)$ be the \textit{conditional density} of $Y$ given $C$ and $V$ evaluated at $c$ and $v$, it follows that (7.31) can be written as
\[ \left[ \exp\{-\alpha(V)\} \int \exp\{-\delta(y, v)\} p_{Y|V,C}(y|v, 1) dy + 1 \right] \pr(C = 1|V = v) = 1, \] (7.32)
where $\pr(C = 1|V)$ is the true probability that $C = 1$ given $V$, which can be modeled and estimated and is thus \textit{identifiable} from the observed data.

Note then that, from (7.32), for fixed $\delta(Y, V)$, $\alpha(V)$ is \textit{identifiable} from the observed data, as it can be written as
\[ \exp\{-\alpha(V)\} = \left\{ \frac{1}{\pr(C = 1|V)} - 1 \right\} / E[\exp\{-\delta(Y, V)\}|V, C = 1], \]
for which the right hand side is identifiable.
Note also that
\[ \mu = E \left\{ \frac{C Y}{\pi(Y, V; \alpha, \delta)} \right\}. \quad (7.33) \]

The foregoing arguments demonstrate that it is thus possible to evaluate sensitivity to the MAR assumption to the specific departure from this assumption embodied in the sensitivity function \( \delta(Y, V) \) using the observed data.

**IMPLEMENTATION:** We discuss how such a sensitivity analysis can be implemented in the specific case of the sensitivity function

\[ \delta(Y, V) = \delta Y \]

in (7.28) and

\[ \alpha(V) = \alpha_0 + \alpha_1^T V \]

in (7.29). Here, the assumed model (7.27) that forms the basis for the sensitivity analysis is

\[ \logit\{pr(C = 1|Y, V)\} = \alpha_0 + \alpha_1^T V + \delta V, \quad \pi(Y, V; \alpha, \delta) = \expit(\alpha_0 + \alpha_1^T V + \delta V). \quad (7.34) \]

For fixed \( \delta \), the strategy is to estimate \( \alpha(\delta) = \{\alpha_0(\delta), \alpha_1(\delta)\}^T \), say, and then use this to estimate \( \mu(\delta) \).

To estimate \( \alpha(\delta) \) under the assumed model (7.34), we take advantage of the fact that (verify)

\[ E \left[ h(V) \left\{ \frac{C}{\pi(Y, V; \alpha, \delta)} - 1 \right\} \right] = 0, \quad (7.35) \]

where \( h(V) \) is a vector of functions of \( V \) whose dimension is the same as the number of unknown parameters in \( \alpha \).

Considering the term in brackets in (7.35) as an unbiased estimating function, an estimator for \( \alpha(\delta) \) for fixed \( \delta \) can be obtained as the solution to the estimating equation

\[ \sum_{i=1}^{N} h(V_i) \left( C_i [1 + \exp\{- (\alpha_0 + \alpha_1^T V_i) + \delta Y_i]\} - 1 \right) = 0. \quad (7.36) \]

A possible choice for \( h(V) \) in (7.36) is to take the partial derivative of \( \alpha(V) \) with respect to the parameters. Here, this leads to

\[ h(V) = \left( \begin{array}{c} 1 \\ V \end{array} \right). \]
Another suggestion follows from the fact that the estimating equation (7.36) can be written as

\[ \sum_{i=1}^{N} \frac{h(V_i)}{\pi(Y_i, V_i; \alpha, \delta)} \{ C_i - \pi(Y_i, V_i; \alpha, \delta) \} = 0. \] (7.37)

Recall that \( \Pr(C = 1 | V) \) can be estimated based on the observed data; this can be accomplished by positing a model

\[ \logit\{\tilde{\pi}(V; \tilde{\alpha})\} = \tilde{\alpha}_0 + \tilde{\alpha}_1 V, \]

which can be fitted by solving the usual score equations for logistic regression; namely,

\[ \sum_{i=1}^{N} \left( \begin{array}{c} 1 \\ V_i \end{array} \right) \{ C_i - \expit(\tilde{\alpha}_0 + \tilde{\alpha}_1 V_i) \} = 0 \]

in \( \tilde{\alpha}_0, \tilde{\alpha}_1 \) to obtain estimators \( \hat{\tilde{\alpha}}_0, \hat{\tilde{\alpha}}_1 \). Then choose

\[ h(V) = \left( \begin{array}{c} 1 \\ V \end{array} \right) \tilde{\pi}(V; \hat{\tilde{\alpha}}). \]

Substituting in (7.37), this choice leads to the estimating equation

\[ \sum_{i=1}^{N} \left( \begin{array}{c} 1 \\ V_i \end{array} \right) \frac{\tilde{\pi}(V_i; \hat{\tilde{\alpha}})}{\pi(Y_i, V_i; \alpha, \delta)} \{ C_i - \pi(Y_i, V_i; \alpha, \delta) \} = 0. \]

This choice has the property that, in the particular case \( \delta = 0 \), corresponding to MAR, the resulting estimator for \( \alpha \) reduces to the MLE, which is efficient.

Once an estimator \( \hat{\alpha}(\delta) = \{ \hat{\alpha}_0(\delta), \hat{\alpha}_1(\delta) \}^T \) is obtained by one of these methods for fixed \( \delta \), the corresponding estimator \( \hat{\mu}(\delta) \) is obtained as, using (7.33), as

\[ \hat{\mu}(\delta) = N^{-1} \sum_{i=1}^{N} \frac{C_i Y_i}{\pi(Y_i, V_i; \hat{\alpha}(\delta), \delta)}. \]

The asymptotic variance of \( \hat{\mu}(\delta) \) can be estimated via the sandwich technique in a manner similar to that described previously, and associated confidence intervals derived.

As in the simpler case in Section 7.1, a sensitivity analysis to the MAR assumption would involve calculating estimates \( \hat{\mu}(\delta) \) and corresponding confidence intervals over a range of \( \delta \) values that includes \( \delta = 0 \) and examining changes in the estimator and its associated confidence interval.
7.3 Longitudinal data with dropout

**SELECTION MODEL APPROACH:** We sketch the main features of a selection model approach to sensitivity analysis in this setting. A pattern mixture approach is described in Chapter 5 of the National Research Council (2010) report; however, there are some errors in the presentation, so we do not consider this here.

Consider the situation in Section 5.3 where longitudinal data \((Y_j, V_j), j = 1, \ldots, T\), on an outcome and auxiliary covariates are to be collected at time points \(t_1 < \cdots < t_T\), where \(t_1\) represents baseline. In addition, a set of baseline covariates \(X\) is collected. The full data are then \(Z = \{(Y_1, V_1), \ldots, (Y_T, V_T), X\}\), and, writing \(Y = (Y_1, \ldots, Y_T)^T\), interest focuses on a semiparametric model

\[
E(Y|X = x) = \mu(x; \beta) = \begin{pmatrix}
\mu_1(x; \beta) \\
\vdots \\
\mu_T(x; \beta)
\end{pmatrix}.
\]

As in Section 5.3, suppose that some individuals drop out, resulting in a monotone missingness pattern. Define \((R_1, \ldots, R_T, R_{T+1})^T\) as usual. Assume that \(X\) is always observed, as is \((Y_1, V_1) (R_1 = 1\) and \(R_{T+1} = 1)\), but, if an individual drops out at time \(t_j, j = 2, \ldots, T\), indicated by the variable \(D = j\), so that s/he is last seen at time \(t_{j-1}\), the history

\[
H_{j-1} = Z_{(j)} = \{X, (Y_1, V_1), \ldots, (Y_{j-1}, V_{j-1})\}
\]

is observed and \{(Y_j, V_j) \ldots, (Y_T, V_T)\} is missing. As before, \(D = T + 1\) corresponds to observing the full data.

Under the MAR assumption, as discussed in Section 5.3, it is assumed that the cause-specific hazard function of dropout given in (5.30),

\[
\lambda_j(Z) = \text{pr}(D = j|D \geq j, Z), \quad j = 2, \ldots, T,
\]

satisfies

\[
\lambda_j(Z) = \text{pr}(D = j|D \geq j, Z) = \text{pr}(D = j|D \geq j, H_{j-1}) = \lambda_j(H_{j-1}), \quad j = 2, \ldots, T,
\]

so that, for individuals who are at risk for dropping out at time \(t_j\), the probability of doing so as a function of the full data depends only on the observed history prior to \(t_j\) and not on observations at \(t_j\) or in the future.
Under MAR, as described in Section 5.3, one would posit *models*

\[ \lambda_j(H_{j-1}; \psi), \]

say, where for each \( j \) dependence on an entire collection of parameters \( \psi \) is likely through a \( j \)-specific subset. An obvious choice would be a *logistic regression* model for each \( j = 2, \ldots, T \), e.g.,

\[ \lambda_j(H_{j-1}; \psi) = \frac{\exp\{\alpha_j(H_{j-1}; \psi)\}}{1 + \exp\{\alpha_j(H_{j-1}; \psi)\}}, \]

or, equivalently, \( \logit\{\lambda_j(H_{j-1}; \psi)\} = \alpha_j(H_{j-1}; \psi) \), for functions \( \alpha_j(H_{j-1}; \psi), j = 2, \ldots, T. \)

A *sensitivity analysis* introduces a *sensitivity function* into the dropout specification. Specifically, we might assume

\[ \logit\{\lambda_j(Z) = \alpha_j(H_{j-1}) + \delta_j(Z), \quad j = 2, \ldots, T, \]  

(7.40)

where \( \delta_j(Z) \) is a *sensitivity function* involving components of \( Z \) beyond those in \( H_{j-1} \). Clearly, \( \delta_j(Z) = 0 \) for all \( j = 2, \ldots, T \) corresponds to the MAR assumption.

As in the previous sections, in specifying the models (7.40), in practice, one would adopt parametric functions \( \alpha_j(H_{j-1}; \psi) \) and choose interpretable *sensitivity functions* for each \( j \) depending on a finite number of parameters. In this setting, choice of the sensitivity functions is *complicated* by the fact that, for each \( j \), possible dependence on components of \( Z \) beyond those in \( H_{j-1} \) could take on many different forms.

One popular approach in this case is to make the assumption of so-called *non future dependence* (NFD). Here, one restricts attention to departures from MAR that arise because the hazards of dropping out at \( t_j, \lambda_j(Z) \), depend only on the history \( H_{j-1} \) and the possibly unobserved values \( (Y_j, V_j) \) at \( t_j \) but not on additional components of \( Z \) in the future; that is

\[ \lambda_j(Z) = \lambda(H_{j-1}, Y_j, V_j), \]  

(7.41)

where this notation emphasizes the dependence on \( (Y_j, V_j) \).

The NFD assumption is of course *not identifiable* from the observed data. It has *practical appeal* in the sense that it seems counterintuitive that data that are not observed in the future should be implicated in whether or not an individual drops out at time \( j \). Of course, it could be that dropout depends on variables *not contained in \( Z \) that are correlated* with future observations, which would render the NFD assumption suspect. We restrict attention here to the NFD assumption, recognizing that it may be an oversimplification.
Accordingly, we consider models (7.40) consistent with the NFD assumption (7.41) of the form
\[
\logit\{\lambda_j(Z)\} = \alpha_j(H_{j-1}; \psi) + \delta_j(H_{j-1}, Y_j, V_j), \quad j = 2, ..., T, \tag{7.42}
\]
so that we can write
\[
\lambda_j(Z) = \lambda_j(H_{j-1}, Y_j, V_j).
\]
In (7.42), then, \(\delta_j(H_{j-1}, Y_j, V_j), j = 2, ..., T\), are fixed sensitivity functions that incorporate the NFD assumption. For example, we might take
\[
\delta_j(H_{j-1}, Y_j, V_j) = \delta Y_j,
\]
in which case dependence on unobserved components of \(Z\) is through the outcome at \(t_j\).

Alternatively, under NFD, a sensitivity function that acknowledges that the dependence of the missingness mechanism on the unobserved \(Y_j\) at \(t_j\) might be different for different values of baseline covariates is
\[
\delta_j(H_{j-1}, Y_j, V_j) = (\delta_0 + \delta^T X) Y_j.
\]
This might be relevant in a clinical trial where \(X\) includes treatment assignment, and there is concern that missingness is differential by treatment.

The parameter of interest is \(\beta\) in the semiparametric model (7.38). In Chapter 5, we discussed the lass of weighted generalized estimating equations that can be solved under MAR to obtain estimators for \(\beta\). We present the basic idea underlying a sensitivity analysis for \(\beta\) by focusing on the general form of augmented inverse probability weighted complete case estimating functions of the form in (5.40) but under the NFD assumption,
\[
\frac{R_T}{\pi_T(H_{T-1}, Y_T, V_T)} \sum_{j=1}^{T} A_j(X)\{Y_j - \mu_j(X; \beta)\} + \sum_{j=1}^{T-1} \left\{ \frac{R_j}{\pi_j(H_{j-1}, Y_j, V_j)} - \frac{R_{j+1}}{\pi_{j+1}(H_j, Y_{j+1}, V_{j+1})} \right\} f_j(H_j). \tag{7.43}
\]
In (7.43),
\[
\pi_j(H_{j-1}, Y_j, V_j) = \prod_{\ell=1}^{j} \{1 - \lambda_\ell(H_{\ell-1}, Y_{\ell}, V_{\ell})\}, \quad j = 2, ..., T.
\]
From (7.42), if we fix the sensitivity functions \(\delta_j(H_{j-1}, Y_j, V_j), j = 2, ..., T\), if we can then estimate \(\psi\) in \(\alpha_j(H_{j-1}; \psi)\) for \(j = 2, ..., T\) from the observed data, we can substitute in (7.43) to obtain estimators \(\beta(\delta)\), say.
Motivated by the estimating equations for $\psi$ in (5.43), consider the estimating equations

$$
\sum_{i=1}^{N} \sum_{j=2}^{T} R_{i,j-1} h_j(H_{i,j-1}; \psi) \left\{ \frac{l(D_i = j)}{\lambda_j(H_{i,j-1}, Y_{ij}, V_{ij}; \psi)} - 1 \right\} = 0,
$$

which can also be written as

$$
\sum_{i=1}^{N} \sum_{j=2}^{T} R_{i,j-1} h_j(H_{i,j-1}; \psi) \left\{ l(D_i = j) \left( 1 + \exp\left\{ -\{\alpha_j(H_{i,j-1}; \psi) + \delta_j(H_{i,j-1}; Y_{ij}, V_{ij})\} \right\} \right\} - 1 \right\} = 0 \quad (7.44)
$$

for functions $h_j(H_{j-1}; \psi)$.

The recommended choice is

$$
h_j(H_{j-1}; \psi) = \frac{\partial}{\partial \psi} \{ \alpha_j(H_{i,j-1}; \psi) \} \lambda_j(H_{j-1}; \hat{\psi}, \delta = 0),
$$

where $\hat{\psi}$ are the MLEs for $\psi$ when $\delta = 0$ found by solving (5.43).

The sensitivity analysis would then proceed in a manner analogous to that described in the previous sections.

### 7.4 Discussion

As noted at the beginning of this chapter, sensitivity analysis is an evolving area. We have reviewed the basic principles underlying popular approaches to sensitivity analysis. Chapter 5 of the National Research Council (2010) report presents examples of graphical displays, in particular plots of sensitivity parameters such as $\Delta$ and $\delta$ against parameter estimates and confidence intervals to provide a visual depiction of sensitivity. Other tactics have been proposed in the literature and were not discussed here.

The National Research Council (2010) report recommends that sensitivity analyses examining departures from an assumed missing data mechanism should be a mandatory part of the reporting of primary findings of clinical trials. The literature on this topic will undoubtedly continue to grow.