# 16 Minimax Theory

## 1 Introduction

Minimax theory has been lurking in the background of much of our work this semester. For example, when estimating Holder-smooth functions with smoothness index $s$ we have tried to obtain rates of the form

$$n^{-1/(2+d/s)},$$

taking for granted that this is the best one can do. But in precisely what sense is this the optimal rate for estimating smooth functions? And how does one prove such optimality results?

Minimax rates are a natural way to characterize optimality. The minimax rate over a model (i.e., set of possible distributions) $\mathcal{P}$ for estimating a particular quantity $\psi_P$ with loss $\ell$ is given by

$$R_n = \inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\{\ell(\widehat{\psi}, \psi_P)\}.$$

In words, this is the best possible (worst-case) estimation error, across all possible estimators.

Minimax rates have crucial implications, both practical and theoretical. First, they give a precise benchmark for the best possible performance of some statistical task. This is useful because, if a given estimator cannot be shown to achieve some lower bound on the minimax rate, then we know we need to work harder – either at improving the estimator (or its analysis), or at improving the lower bound. Conversely, if a given estimator does achieve the minimax rate, then we know it cannot be improved (at least in terms of rates) without adding or changing assumptions. Second, minimax rates precisely characterize the fundamental limits of estimation in a given problem. This allows us to compare different tasks and order them in terms of their inherent statistical difficulty.

Minimax rates are well-understood in many problems (but certainly not all). Consider mean absolute error for example. Then for smooth nonparametric regression/density estimation the minimax rate is

$$n^{-1/(2+d/s)}.$$

This is like the parametric rate $1/\sqrt{n}$ but with an added penalty for how large the dimension is relative to the smoothness. For smooth functional estimation (e.g., $\psi = \mathbb{E}\{p(Z)\} = \int p(z)^2\, dz$ the expected density), the minimax rate is

$$\max\left\{n^{-1/\left(1+\frac{d}{4s}\right)}, 1/\sqrt{n}\right\}.$$

Therefore functional estimation is in this sense easier than regression: the rate is parametric whenever $s \geq d/4$, and otherwise the penalty driving the rate away from parametric is weaker than for regression, depending on $d/4s$ instead of $d/s$. For smooth density estimation with Gaussian measurement error, the minimax rate is

$$(\log n)^{-s/2}$$

which is very slow, indicating the inherent difficulty of the problem. This illustrates that in this setting one either needs to accept slow rates or else add some structure/assumptions.

## 2 Reduction to hypothesis testing

At a high level, deriving minimax rates requires two separate steps. First, one upper bounds the minimax rate up to constants, i.e., proves that for some rate $r_n$ we have

$$\inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\{\ell(\widehat{\psi}, \psi_P)\} \equiv R_n \leq C r_n.$$

This is accomplished by constructing a specific estimator $\widehat{\psi}$ and proving that it satisfies

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P\{\ell(\widehat{\psi}, \psi_P)\} \leq C r_n,$$

since the risk for a particular estimator is certainly no smaller than the minimax rate. Second, one lower bounds the minimax rate

$$R_n \geq C' r_n.$$

This is the step we focus on here, since we have already seen many upper bound examples.

Tsybakov [2009] describes three reductions or simplications that are generally useful in thinking about and deriving minimax lower bounds.

1. The first reduction is to minimax probabilities. By Markov's inequality, we have

   $$\mathbb{P}(\ell(\widehat{\psi}, \psi) \geq s) \leq \mathbb{E}\{\ell(\widehat{\psi}, \psi)\}/s$$

   so that

   $$\inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\{\ell(\widehat{\psi}, \psi_P)\} \geq s \inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} P(\ell(\widehat{\psi}, \psi_P) \geq s).$$

2. The second reduction is to a finite set of hypotheses. Clearly we have

   $$\inf_{\widehat{\psi}} \sup_{P} P(\ell(\widehat{\psi}, \psi_P) \geq s) \geq \inf_{\widehat{\psi}} \max_{P_0, P_1, \ldots, P_M} P_j(\ell(\widehat{\psi}, \psi_j) \geq s)$$

   for any set $(P_1, ..., P_M) \subseteq \mathcal{P}$.

3. The third reduction is to minimax hypothesis testing via separated hypotheses. If

   $$\ell(\psi_j, \psi_k) \geq s, \ \forall k \neq j$$

   then for any $\widehat{\psi}$ we have

   $$P_j(\ell(\widehat{\psi}, \psi_j) \geq s/2) \geq P_j(T^* \neq j)$$

   where $T^*$ is the minimum distance test defined as $T^* = \arg\min_k \ell(\widehat{\psi}, \psi_k)$. This follows since if $T^* \neq j$ then taking $k = T^*$ we have by the triangle inequality that

   $$s \leq \ell(\psi_j, \psi_k) \leq \ell(\psi_j, \widehat{\psi}) + \ell(\widehat{\psi}, \psi_k) \leq 2\ell(\widehat{\psi}, \psi_j)$$

   and so $T^* \neq j$ implies $\ell(\widehat{\psi}, \psi_j) \geq s/2$.

Putting the above reductions together, we have

$$\inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\{\ell(\widehat{\psi}, \psi_P)\} \geq \left(\frac{s}{2}\right) \inf_T \max_{0 \leq j \leq M} P_j(T \neq j) \equiv \left(\frac{s}{2}\right) p_{e,M}$$

where $\inf_T$ denotes the infimum over all tests (i.e., all functions of the data that return a guess at the index $j$ of the true $\psi_j$). The quantity $p_{e,M}$ is referred to as the minimax probability of error for testing hypotheses $P_0, P_1, ..., P_M$ (or just the minimax testing rate), and it can be lower bounded for example using the Neyman-Pearson lemma, which implies that maximum likelihood tests have minimal error for simple versus simple hypothesis testing.

## 3 Distributional distances

Lower bounds on the minimax testing rate depend crucially on how close the distributions $P_j$ are to each other. Therefore we need some basic information about distributional distances. In what follows we let $\nu$ denote a dominating measure for distributions $P$ and $Q$.

$$H(P,Q) = \sqrt{\int (\sqrt{p} - \sqrt{q})^2 \, d\nu} \qquad \text{Hellinger}$$

$$TV(P,Q) = \sup_{A \in \mathcal{A}} \left| \int_A (p - q) \, d\nu \right| \qquad \text{Total variation}$$

$$\|P - Q\|_1 = \int |p - q| \, d\nu = 2TV(P,Q) \qquad L_1$$

$$KL(P,Q) = \int \log\left(\frac{dP}{dQ}\right) \, dP \qquad \text{Kullback-Leibler}$$

$$\chi^2(P,Q) = \int \frac{(p-q)^2}{q} \, d\nu \qquad \chi^2$$

There are numerous ways to link these different distances, which are all examples of $f$-divergences $D_f(P\|Q) = \int f\left(\frac{dP}{dQ}\right) \, dQ$ for some choice of $f$. For example,

$$\frac{1}{2}H^2(P,Q) \leq TV(P,Q) \leq \sqrt{KL(P,Q)} \leq \sqrt{\chi^2(P,Q)}.$$

There are also some useful expressions for distances between two product measures $P^n = \otimes_{i=1}^n P_i$ and $Q^n = \otimes_{i=1}^n Q_i$, such as

$$H^2(P^n, Q^n) = 2\left(1 - \prod_{i=1}^n \left(1 - \frac{H^2(P_i, Q_i)}{2}\right)\right)$$

$$KL(P^n, Q^n) = \sum_{i=1}^n KL(P_i, Q_i)$$

$$\chi^2(P^n, Q^n) = \prod_{i=1}^n \Big(1 + \chi^2(P_i, Q_i)\Big) - 1.$$

See Section 2.4 of Tsybakov [2009] for other results and more details.

# 4 Le Cam's two-point method

Heuristically, minimax lower bounds are essentially proved by carefully constructing two distributions that are so close that they are statistically indistinguishable (i.e., we would not have power to test whether observations were coming from one distribution versus another), but for which the parameter of interest $\psi$ is separated as much as possible across the distributions. Then a lower bound arises via proof by contradiction: suppose there is an estimator whose risk is smaller than the separation in $\psi$, then we would be able to distinguish the two distributions. But this is a contradiction, since we specifically constructed the distributions so that they were indistinguishable; therefore no estimator with risk smaller than the separation exists.

A version of this argument based on $\chi^2$ distances is formalized in the following result. Minimax lower bounds based on two distributions in this way are referred to Le Cam's two-point method; this stems from Le Cam (1973), which mostly considered parametric problems.

**Lemma 1.** *For $P_0, P_1 \in \mathcal{P}$, let $\psi(P_j) = \psi_j$. If $\ell(\psi_1, \psi_0) \geq s > 0$ and $\chi^2(P_1, P_0) \leq \alpha < \infty$ then*

$$\inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\Big\{\ell(\widehat{\psi}, \psi_P)\Big\} \;\geq\; \left(\frac{s}{8}\right)\exp(-\alpha).$$

*Proof.* From the reductions in Section 2, the minimax rate is lower bounded as $R_n \geq (s/2)p_{e,M}$. Now

$$\begin{aligned} p_{e,M} \equiv \inf_T \max_{j=0,1} P_j(T \neq j) &\geq \inf_T \frac{P_0(T \neq 0) + P_1(T \neq 1)}{2} \\ &\geq \frac{1}{2}\int \min(dP_1, dP_0) = (1 - TV(P_0, P_1))/2 \end{aligned}$$

where the first line follows since maxima are greater than averages, and the second since the error is minimized by the maximum likelihood test $T = \mathbb{1}(dP_1 \geq dP_0)$, whose error is given by the expression given. Lemma 2.6 of Tsybakov [2009] then shows $\int \min(dP_1, dP_0) \geq \exp(-KL(P_1, P_0))/2$, which gives the result since $KL(P_1, P_0) \leq \chi^2(P_1, P_0)$. □

*Remark* 1. Lemma 1 uses the $\chi^2$ distance, but analogous results can be proved for Hellinger, total variation, and Kullback-Leibler distances as well. See Theorem 2.2 of Tsybakov [2009].

**Theorem 1.** *Let $\mathcal{P}$ denote the model where $\mu(x) = \mathbb{E}(Y \mid X = x)$ is $s$-smooth, and $X$ has a bounded density. Then the minimax rate for estimating $\mu(x_0)$ is lower bounded as*

$$\inf_{\widehat{\mu}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\Big\{|\widehat{\mu}(x_0) - \mu(x_0)|\Big\} \geq C n^{-1/(2+d/s)}$$

*for some constant $C$ depending on $s$ and $d$ but not $n$.*

*Proof.* Let $Y \in \{0,1\}$ and $x_0 = 1/2$ without loss of generality, and consider the distributions

$$p_0(y,x) = \mathbb{1}\left(x \in [0,1]^d\right) 1/2$$
$$p_1(y,x) = \mathbb{1}\left(x \in [0,1]^d\right) \left\{1/2 + (2y-1)h^s B\left(\frac{x-1/2}{h}\right)\right\},$$

where $X \sim \text{Unif}(0,1)^d$ and the regression function $\mathbb{E}(Y \mid X = x)$ is constant under $P_0$ and has a small bump at $x = 1/2$ under $P_1$. We assume the bump function $B$ is sufficiently smooth and has bounded support. See Equation 2.34 of Tsybakov [2009] for examples.

The first task is to show that $P_0, P_1 \in \mathcal{P}$, which really amounts to showing that the regression function under $P_1$ is $s$-smooth. This is proved in Equation 2.35 of Tsybakov [2009].

At $x_0 = 1/2$ the separation in $\mu(x)$ under $P_1$ versus $P_0$ is $B(0)h^s \equiv C_1 h^s$. Thus now we need to compute the distance between $P_1$ and $P_0$; we use $\chi^2$ distance. Note we have

$$\chi^2(P_1,P_0) = \sum_y \int_{\mathcal{X}} \frac{\{p_1(y,x) - p_0(y,x)\}^2}{p_0(y,x)}\, dx = \sum_y \int 2\left\{(2y-1)h^s B\left(\frac{x-1/2}{h}\right)\right\}^2 dx$$
$$= 2h^{2s}\int B\left(\frac{x-1/2}{h}\right)^2 dx = 2h^{2s+d}\int B(u)^2\, du \equiv C_2 h^{2s+d}$$

Therefore

$$\chi^2(P_1^n, P_0^n) = \prod_{i=1}^n \left(1 + \chi^2(P_1,P_0)\right) - 1$$
$$= \left(1 + C_2 h^{2s+d}\right)^n - 1 \le \exp\left(C_2 n h^{2s+d}\right).$$

Now for some $1 < \alpha < \infty$ we take $\exp(C_2 n h^{2s+d}) = \alpha$, i.e., $h = (\log\alpha / C_2 n)^{1/(2s+d)}$, so that the $\chi^2$ distance is bounded above by $\alpha$ and the separation in $\mu(x_0)$ under $P_1$ versus $P_0$ is $C_1 h^s = C_1 (\log\alpha/C_2 n)^{s/(2s+d)} \equiv C_3 n^{-1/(2+d/s)}$. Applying Lemma 1 now yields the result. □

## 5 Fano's lemma

Le Cam's method does not give tight bounds for $L_p$ distances, e.g., $\mathbb{E}\|\widehat{\mu} - \mu\|_p$. For this we need an approach based on Fano's inequality, illustrated in the following result.

**Lemma 2.** *Let $P_0, P_1, ..., P_M \in \mathcal{P}$ with $\psi(P_j) = \psi_j$ and assume:*

1. *$\ell(\psi_j, \psi_k) \ge s > 0$ for all $j < k$,*
2. *$\frac{1}{M}\sum_{j=1}^M \chi^2(P_j, P_0) \le \alpha M$ for some $0 < \alpha < 1/2$.*

*Then*

$$\inf_{\widehat{\psi}} \sup_{P\in\mathcal{P}} \mathbb{E}_P\Big\{\ell(\widehat{\psi}, \psi_P)\Big\} \;\ge\; \left(\frac{s}{4}\right)\left(1 - \alpha - \frac{1}{M}\right).$$

Other tools for deriving minimax lower bounds include Assouad's lemma and the van Trees inequality. See 2.7.2–2.73 of Tsybakov [2009] for details.

## 6 Fuzzy hypotheses

For hypothesis testing and functional estimation problems, it is not enough to construct lower bounds based on two distributions or even $M \geq 2$ distributions via Fano's method. Instead, to obtain tight lower bounds in these kinds of problems one needs to carefully construct prior distributions over the model, i.e., "fuzzy" mixtures of distributions. The following result shows how to obtain lower bounds in such cases.

**Lemma 3.** *Let $\varpi_0$ and $\varpi_1$ denote two prior distributions over $\mathcal{P}$, which induce $n$-fold product distributions $Q_j(E) = \int P^n(E)\, d\varpi_j(P)$ for measurable set $E$. If for a functional $\psi : \mathcal{P} \to \mathbb{R}$ we have*

$$\varpi_0(\{P : \psi(P) \leq c\}) \geq 1 - \beta_0$$
$$\varpi_1(\{P : \psi(P) \geq c + s\}) \geq 1 - \beta_1$$

*and also*

$$H^2(Q_0, Q_1) \leq \alpha < 2$$

*then*

$$\inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\left(|\widehat{\psi} - \psi_P|\right) \;\geq\; \left(\frac{s}{2}\right)\left(\frac{1 - \sqrt{\alpha(1-\alpha/4)}}{2} - \frac{\beta_0 + \beta_1}{2}\right).$$

Just as before, Lemma 3 shows how, for a good lower bound, one must carefully construct two prior distributions that simultaneously ensure: (1) enough separation of the parameter of interest, but also (2) enough similarity so that the Hellinger distance is appropriately small.

**Theorem 2.** *Let $\mathcal{P}$ denote the model where $\mu(x) = \mathbb{E}(Y \mid X = x)$ is $s$-smooth, and $X$ has a bounded density. Then the minimax rate for estimating $\psi = \mathbb{E}\{\mu(X)^2\}$ is lower bounded as*

$$\inf_{\widehat{\psi}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\left(|\widehat{\psi} - \psi_P|\right) \geq C\left(n^{-1/(1+d/4s)} + \frac{1}{\sqrt{n}}\right)$$

*for some constant $C$ depending on $s$ and $d$ but not $n$.*

*Proof.* The $C/\sqrt{n}$ lower bound can be derived using Le Cam's two-point method. Therefore we focus on the other term, corresponding to slower rates.

First let $B$ denote a bump function that is sufficiently smooth with bounded support (cf. Eq. 2.34 of Tsybakov [2009]), and for which $\int B(u)\, du = 0$. Also let $(\mathcal{X}_1, ..., \mathcal{X}_k)$ denote a partition of the support of $X$ into $k$ cubes of equal size (for $k$ an integer raised to the power $d$) with midpoints $m_j$, $j = 1, ..., k$, so each cube $\mathcal{X}_j$ has side length $1/k^{1/d}$. Now let $Y \in \{0, 1\}$, and consider the distributions

$$p_0(y, x) = \mathbb{1}\left(x \in [0,1]^d\right) 1/2$$
$$p_\lambda(y, x) = \mathbb{1}\left(x \in [0,1]^d\right)\left\{1/2 + (2y-1)k^{-s/d}\sum_{j=1}^{k} \lambda_j B\left(\frac{x - m_j}{k^{-1/d}}\right)\right\},$$

where $\lambda_j \in \{0,1\}$. Under $P_0$ we use a point mass prior $\varpi_0$ that puts all of its mass on $p_0$; under $P_1$ we use a prior $\varpi_1$ that puts equal mass on each $\lambda = (\lambda_1, ..., \lambda_k)$, so that the $\lambda_j$ can be viewed as iid Rademacher random variables.

To show that $P_0, P_1 \in \mathcal{P}$, i.e., that the regression function under $P_\lambda$ is $s$-smooth, one can use the same logic as in Equation 2.35 of Tsybakov [2009].

For the separation of $\psi_P = \mathbb{E}_P\{\mu_P(X)^2\}$ under $P_0$ versus $P_\lambda$ note that $\psi_0 = \frac{1}{4}$ and

$$\begin{aligned}
\psi_\lambda &= \int \left\{ 1/2 + k^{-s/d} \sum_{j=1}^{k} \lambda_j B\left(\frac{x - m_j}{k^{-1/d}}\right) \right\}^2 \, dx \\
&= \frac{1}{4} + k^{-2s/d} \sum_{j=1}^{k} \sum_{\ell=1}^{k} \lambda_j \lambda_\ell \int B\left(\frac{x - m_j}{k^{-1/d}}\right) B\left(\frac{x - m_\ell}{k^{-1/d}}\right) \, dx \\
&= \frac{1}{4} + k^{-2s/d} \sum_{j=1}^{k} \int_{\mathcal{X}_j} B\left(\frac{x - m_j}{k^{-1/d}}\right)^2 \, dx \\
&= \frac{1}{4} + k^{-2s/d} \frac{1}{k} \sum_{j=1}^{k} \int B\left(u\right)^2 \, dx \\
&= \frac{1}{4} + C_1 k^{-2s/d}
\end{aligned}$$

where in the second line we used the fact that $\int B(u)\, du = 0$ and that $B$ has bounded support (so that products of bumps on different midpoints are zero). Therefore in Lemma 3 we can take $\beta_0 = \beta_1 = 0$ with $c = 1/4$ and $s = C_1 k^{-2s/d}$.

For the Hellinger distance, we use the following lemma from Robins et al. [2009], which helpfully relates the distance between the $n$-fold product distributions to those of simpler posteriors over a single observation.

**Lemma 4.** *Let $P_\lambda$ and $Q_\lambda$ denote distributions indexed by $\lambda = (\lambda_1, \ldots, \lambda_k)$, and let $\mathcal{Z} = \cup_{j=1}^{k} \mathcal{Z}_j$ denote a partition of the sample space. Assume $P_\lambda(\mathcal{Z}_j) = Q_\lambda(\mathcal{Z}_j) = p_j$ for all $\lambda$, and the conditional distributions $\mathbb{1}_{\mathcal{Z}_j} dP_\lambda / p_j$ and $\mathbb{1}_{\mathcal{Z}_j} dQ_\lambda / p_j$ do not depend on $\lambda_\ell$ for $\ell \neq j$. For a prior distribution $\varpi$ over $\lambda$, let $\overline{p} = \int p_\lambda \, d\varpi(\lambda)$ and $\overline{q} = \int q_\lambda \, d\varpi(\lambda)$, and define*

$$\begin{aligned}
\delta_1 &= \max_{j \in S} \sup_{\lambda} \int_{\mathcal{Z}_j} \frac{(p_\lambda - \overline{p})^2}{p_\lambda p_j} \, d\nu \\
\delta_2 &= \max_{j \in S} \sup_{\lambda} \int_{\mathcal{Z}_j} \frac{(q_\lambda - p_\lambda)^2}{p_\lambda p_j} \, d\nu \\
\delta_3 &= \max_{j \in S} \sup_{\lambda} \int_{\mathcal{Z}_j} \frac{(\overline{q} - \overline{p})^2}{p_\lambda p_j} \, d\nu
\end{aligned}$$

*for a dominating measure $\nu$. If $\overline{p}/p_\lambda \leq b < \infty$ and $n p_j \max(1, \delta_1, \delta_2) \leq b$ for all $j$, then*

$$H^2\left( \int P_\lambda^n \, d\varpi(\lambda), \int Q_\lambda^n \, d\varpi(\lambda) \right) \leq Cn \left\{ n \left( \max_j p_j \right) \left( \delta_1 \delta_2 + \delta_2^2 \right) + \delta_3 \right\}$$

*for a constant $C$ only depending on $b$.*

In our problem $p_\lambda = \overline{p}$ since $p_0$ is a point null, therefore $\delta_1 = 0$. We also have $\overline{p} = \overline{q}$ so that $\delta_3 = 0$; therefore the only term in the bound is $\delta_2$. For it we have

$$\begin{aligned}
\delta_2 &= \max_{j\in S} \sup_\lambda \int_{\mathcal{Z}_j} \frac{(q_\lambda - p_\lambda)^2}{p_\lambda p_j}\, d\nu \\
&= 2k \max_{j\in S} \sup_\lambda \int_{\mathcal{X}_j} k^{-2s/d} \left\{ \sum_{\ell=1}^{k} \lambda_\ell B\left(\frac{x-m_\ell}{k^{-1/d}}\right)\right\}^2 dx \\
&= 2 \max_{j\in S} k^{-2s/d} k \int_{\mathcal{X}_j} \sum_{\ell=1}^{k} B\left(\frac{x-m_\ell}{k^{-1/d}}\right)^2 dx \\
&= 2 \max_{j\in S} k^{-2s/d} k \int_{\mathcal{X}_j} B\left(\frac{x-m_j}{k^{-1/d}}\right)^2 dx \\
&= 2k^{-2s/d} \int B(u)^2\ du \equiv C_2 k^{-2s/d}.
\end{aligned}$$

Therefore by Lemma 4 since $p_j = 1/k$ we have

$$H^2\left(\int P_\lambda^n\, d\varpi(\lambda), \int Q_\lambda^n\, d\varpi(\lambda)\right) \ \le\ \left(\frac{Cn^2}{k}\right) C_2^2 k^{-4s/d}.$$

Taking $(Cn^2/k)C_2^2 k^{-4s/d} = \alpha$ yields

$$\frac{n^2}{k^{(4s+d)/d}} = \alpha/(CC_2^2) \implies k^{(4s+d)/d} = C_3 n^2 \implies k = (C_3 n^2)^{d/(4s+d)} \equiv C_4 n^{2d/(4s+d)}$$

so that the functional separation is

$$s = C_1 k^{-2s/d} = C_1 \left(C_4 n^{2d/(4s+d)}\right)^{-2s/d} \equiv C_5 n^{-4s/(4s+d)}.$$

Lemma 3 now yields the result. □

## References


J. M. Robins, E. J. Tchetgen Tchetgen, L. Li, and A. W. van der Vaart. Semiparametric minimax rates. *Electronic Journal of Statistics*, 3:1305–1321, 2009.

A. B. Tsybakov. *Introduction to Nonparametric Estimation.* New York: Springer, 2009.