Machine Learning Reading Group Presenter: Aaron Geelon So
Winter 2020, January 27

Generalization in adaptive data analysis

Today, we will present *A new analysis of differential privacy's generalization guarantees* (JLNRSS 2019). It is concerned with the setting of **adaptive data analysis**, where we would like to collect some data $S \sim \mathcal{P}^n$ and adaptively ask a sequence of questions $q_1, \dots, q_T$ about the data. That is, the question q_{t+1} that we ask next could depend on our previous queries and answers we obtained by interacting with S:

$$q_{t+1} \leftarrow \mathcal{A}\big((q_1, a_1), \dots, (q_t, a_t)\big).$$

Here, $\mathcal{A}$ is the data analyst choosing the next query.

For simplicity, let's consider only **linear queries** $q_t : \mathcal{X} \to [0, 1]$ where we aim to determine:

$$q(\mathcal{P}) := \mathop{\mathbb{E}}_{x \sim \mathcal{P}} [q(x)].$$

Notice that if the queries were not adaptive, then to obtain ε-accuracy for all T queries, a Hoeffding-Chernoff concentration bound tells us that we need around $n = \frac{1}{\varepsilon^2} \log T$ samples. However, when the queries are chosen adaptively, the samples and the queries are no longer independent. A naive way to overcome this is to simply resample data for each query, so we need at most $n = \frac{1}{\varepsilon^2} \cdot T$ samples. Thus, the number of samples grows linearly in the adaptive case, compared to the logarithmic growth in the non-adaptive setting.

It turns out that there is a lower bound on the adaptive rate: $\sqrt{T}$. The difficulty of the adaptive setting is that if we learn too much about S, we might be able to find a query q where an answer $q(S)$ on S is very different from $q(\mathcal{P})$. One way to achieve this optimal rate asymptotically is by using a *differentially private mechanism* to perform the adaptive analysis: by perturbing the answers with noise, we prevent the analyst from learning too much about the particular sample S. while hopefully they are still able to learn something nontrivial about the overall underlying distribution $\mathcal{P}$.

The idea here is that **generalization** can be achieved through **algorithmic stability**: if the queries and answers that we get cannot depend too much on small changes in the samples S that we draw, then these answers will likely generalize to the rest of the distribution. Theorems that allow one to convert a differential privacy guarantee into a generalization guarantee are called **transfer theorems**, and we'll show a transfer theorem with an elementary proof, which also illuminates this intuition further.

1 Intuition

1.1 Posterior as an object to describe generalizability

To motivate the technique presented in this paper, consider drawing a sample $S \sim \mathcal{P}^n$ and from it, producing a transcript of our interaction with it, $\pi = \big((q_1, a_1), \dots, (q_T, a_T)\big)$. Then, we could define $\mathcal{Q}_\pi$ to be the *posterior distribution* on samples $S' \in \mathcal{X}^n$ that produces the transcript π,

$$\mathcal{Q}_\pi(S') = \mathcal{P}(S' | \Pi = \pi).$$

In particular, if the mechanism by which we choose queries and answers is deterministic, then $\mathcal{Q}_\pi$ is supported on exactly the samples $S' \in \mathcal{X}^n$ that yield this sequence of answers. Notice that in this case, then we have a uniform bound over the queries we performed:

$$\max_{t \in [T]} \big|a_t - q_t(S)\big| = 0 \quad \Rightarrow \quad \max_{t \in [T]} \big|a_t - q_t(\mathcal{Q}_{\pi(S)})\big| = 0.$$

That is, the answer that we get generalize quite well if it were the case that our samples S were drawn from the posterior distribution $\mathcal{Q}_\pi$ and not $\mathcal{P}^n$. But it's likely here that $\mathcal{Q}_\pi$ and $\mathcal{P}^n$ are very different. And so, just because our answers were close (in fact equal) to $q(\mathcal{Q}_\pi)$, they might be far from $q(\mathcal{P}^n)$.

Intuitively, if someone who looks at our transcript could recover S, then we are very likely to have overfit to our data. Perhaps one measure of how much of our analysis is an artefact of our sample (rather than the population) is the following:

$$\sup_{q\,:\,\mathcal{X}\to[0,1]} \big|q(\mathcal{P}^n) - q(\mathcal{Q}_{\pi(S)})\big|.$$

This, of course, is just the total variation $d_{\mathrm{TV}}(\mathcal{P}^n, \mathcal{Q}_{\pi(S)})$. Under this metric, one extreme way we could prevent overfitting is by neglecting to look at S at all. Then, $\Pi(S)$ is independent to S, and our answers with respect to S is just as bad (or good) as with respect to $\mathcal{P}$. Indeed, $\mathcal{Q}_\pi = \mathcal{P}^n$, which is great for generalization, but this way of minimizing total variation is no good for accuracy.

But perhaps we don't really care if $\mathcal{Q}_\pi$ and $\mathcal{P}^n$ are very different. After all, we care only about the queries we ended up asking, $q_1, \ldots, q_T$. Perhaps we should consider:

$$\max_{t\in[T]} \big|q_t(\mathcal{P}^n) - q_t(\mathcal{Q}_{\pi(S)})\big|.$$

We will use this later to define the notion of **posterior insensitivity**. One way we can decrease this sensitivity is by adding uncertainty to the outcome of the analysis $\pi(S)$. Then, the posterior $\mathcal{Q}_{\pi(S)}$ will reveal less about S, pushing the posterior closer to $\mathcal{P}^n$ (this will be how differential privacy can lead to posterior insensitivity). Though, one consequence may be the degradation of the **in-sample accuracy**:

$$\max_{t\in[T]} \big|a_t - q_t(S)\big|.$$

Having made this trade off, the in-sample accuracy may no longer be zero. But hopefully, we can still prove something like:

$$\max_{t\in[T]} \big|a_t - q_t(S)\big| < \varepsilon \quad \Rightarrow \quad \max_{t\in[T]} \big|a_t - q_t(\mathcal{Q}_{\pi(S)})\big| < \varepsilon'. \tag{1}$$

In which case, the simple application of the triangle inequality would imply that our answers a_t are not only close to $q_t(\mathcal{Q}_{\pi(S)})$, but close to $q_t(\mathcal{P}^n)$. This would mean that if our mechanism by which we obtain answers to our queries is posterior insensitive and in-sample accurate, then this mechanism would also have out-of-sample accuracy over the underlying distribution.

1.2 The brain-damaged statistician

It seems very plausible that we should be able to prove something like Equation 1 in probability; to do this, we need the **Bayesian resampling lemma**. Consider a brain-damaged statistician, who collects some data S then performs some analysis using S to obtain Π. Then, accidentally throwing away S, they decide to sample S' from the posterior given Π. Then, what is the relationship between the distributions of (S, Π) and (S', Π)? It turns out that they are the same. In particular:

$$P(s)P(\pi|s) = P(s,\pi) = \sum_{\tilde{s}} P(\tilde{s})P(\pi|\tilde{s})P(s|\pi) = P(\pi)P(s|\pi) = P(\pi)Q_\pi(s).$$

As a result, this means that the following are equal:

$$\Pr_{S\sim\mathcal{P}^n,\pi\sim\Pi(S)}\left[\max_{t\in[T]} \big|a_t - q_t(S)\big| \geq \varepsilon\right] = \Pr_{S\sim\mathcal{P}^n,\pi\sim\Pi(S),S'\sim\mathcal{Q}_\pi}\left[\max_{t\in[T]} \big|a_t - q_t(S')\big| \geq \varepsilon\right]$$

Then, to go from a bound with respect to an individual draw $S' \sim \mathcal{Q}_\pi$ to a bound with respect to the expectation over $\mathcal{Q}_\pi$, we can use, for example, a Markov or Chernoff bound.

1.3 Differential privacy implies posterior insensitivity

One way to obtain a statistical estimation mechanism that is posterior insensitive is to ensure that it is **differentially private**. In differential privacy, the outcome of the data analysis Π cannot depend too much on any single data point $X_i \in S$. That is, the distribution on Π conditioned on knowing X_i can't be very different from the distribution before this knowledge:

$$d_{\mathrm{TV}}\big((\Pi|X_i = x)\,,\,\Pi\big) \approx 0.$$

On the other hand, posterior insensitivity has to do with how well we can distinguish between the distributions $\mathcal{P}^n$ and $\mathcal{Q}_{\pi(S)}$ using the queries $q_1, \ldots, q_T$. If it is the case that there is some q_t that is able to distinguish these distributions well, that implies that $d_{\mathrm{TV}}\big((\mathcal{P}^n|\Pi), (\mathcal{Q}_\pi|\Pi)\big) \gg 0$. But then, if we choose $X \sim S$ randomly from our sample S, then we must have:

$$d_{\mathrm{TV}}\big((X|\Pi = \pi)\,,\,X\big) \gg 0.$$

But how can it be the case that Π doesn't depend very much on any single data point X, but then knowledge of Π also tells us a lot about one of the data points used in its analysis? And so, we'll show later on that if our mechanism is differentially private, then it must be posterior insensitive.

2 Formal setting

Let $\mathcal{P}$ be a distribution over $\mathcal{X}$ a data domain. A sample $S = (X_1, \ldots, X_n) \overset{\text{i.i.d.}}{\sim} \mathcal{P}^n$ is drawn i.i.d. from $\mathcal{P}$. As noted before, we are concerned with linear queries $q : \mathcal{X} \to [0, 1]$, where we let:

$$q(\mathcal{D}) := \mathop{\mathbb{E}}_{x \sim \mathcal{D}}\big[q(x)\big],$$

for any distribution $\mathcal{D}$ over $\mathcal{X}$. If S is a sample, let us denote the empirical mean by:

$$q(S) := \frac{1}{|S|}\sum_{x \in S} q(x).$$

In the adaptive data analysis setting, given a family Q of linear queries, we have:

- a **statistical estimator** $M : \mathcal{X}^n \times Q^* \to \mathbb{R}^*$, which is a (possibly stateful and randomized) algorithm that interactively returns answers $a_t \in \mathbb{R}$ to queries $q_t \in Q$ using a data set $S \in \mathcal{X}^n$,

$$a_t \leftarrow M(S, q_t)$$

- a **data analyst** $\mathcal{A} : \mathbb{R}^* \to Q^*$, which is any (randomized) algorithm that selects the next query q_{t+1} based on the sequence of previous answers,

$$q_{t+1} \leftarrow \mathcal{A}(a_1, \ldots, a_t)$$

- a **transcript** π of the interaction between the statistical estimator and data analyst,

$$\pi := \big((q_1, a_1), \ldots, (q_T, a_T)\big).$$

  We'll denote the random variable for the transcript by $\Pi(M, \mathcal{A}, S)$, or more succinctly, $\Pi(S)$ if M and $\mathcal{A}$ are clear from context. And as an abuse of notation, we also say $(q_t, a_t) \in \pi$ for $t \in [T]$.

We also denote the **posterior distribution** on dataset conditional on $\Pi = \pi$ by:

$$\mathcal{Q}_\pi = \big(\mathcal{P}^n \,\big|\, \Pi(M, \mathcal{A}, S) = \pi\big).$$

Definition 1 ((α,β)-sample accuracy)**.** A statistical estimator M is (α,β)-*sample accurate* if for every data analyst $\mathcal{A}$ and every data distribution $\mathcal{P}$,

$$\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(M,\mathcal{A},S)}\left[\max_{(q,a)\in\pi}\left|q(S)-a\right|\geq\alpha\right]\leq\beta.$$

For example, the statistical estimator that simply returns the empirical estimate of a query q on S is $(0,0)$-sample accurate. But as we discussed, in-sample accuracy does not imply good generalization, especially in adaptive data analysis. So, we define:

Definition 2 ((α,β)-distributional accuracy)**.** A statistical estimator M is (α,β)-*distributional accurate* if for every data analyst $\mathcal{A}$ and every data distribution $\mathcal{P}$,

$$\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(M,\mathcal{A},S)}\left[\max_{(q,a)\in\pi}\left|q(\mathcal{P}^n)-a\right|\geq\alpha\right]\leq\beta.$$

For example, the statistical estimator that partitions S into T subsets obtains, by Hoeffding's inequality, $\left(\varepsilon,2\exp(-2n\varepsilon^2/T)\right)$-distributional accuracy. And finally, motivated by our earlier discussion on posterior insensitivity, we define it formally as:

Definition 3 ((ε,δ)-posterior insensitive)**.** The interaction $(M,\mathcal{A})$ is called (ε,δ)-*posterior insensitive* if for every data distribution $\mathcal{P}$,

$$\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(M,\mathcal{A},S)}\left[\max_{(q,a)\in\pi}\left|q(\mathcal{P}^n)-q(\mathcal{Q}_\pi)\right|\geq\varepsilon\right]\leq\delta.$$

3 A general transfer theorem

Now, the main transfer theorem:

Theorem 4 (General transfer theorem)**.** *Suppose that M is an (α,β)-sample accurate statistical estimator, and that the interaction $(M,\mathcal{A})$ is (ε,δ)-posterior insensitive. Then, it is (α',β')-distributionally accurate for all $\alpha'=\alpha+c+\varepsilon$ and $\beta'=\frac{\beta}{c}+\delta$ where $c>0$.*

Recall our intuition that if M is (α,β)-sample accurate, then all the answers it produce a_t is close to $q_t(S)$. This should imply that a_t is close to $q_t(\mathcal{Q}_\pi)$. And if furthermore, M is (ε,δ)-posterior insensitive, then $q_t(\mathcal{Q}_\pi)$ is close to $q_t(\mathcal{P})$. Then, by the triangle inequality, this theorem would follow. We need:

Lemma 5 (Sample accuracy implies posterior distributional accuracy)**.** *Suppose that M is (α,β)-sample accurate. Then, it is (α',β')-distributionally accurate over $\mathcal{Q}_\pi$, where $\alpha'=\alpha+c$ and $\beta'=\beta/c$ for all $c>0$. That is,*

$$\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(S)}\left[\max_{(q,a)\in\pi}\left|a-q(\mathcal{Q}_\pi)\right|>\alpha+c\right]\leq\frac{\beta}{c}.$$

Proof of Theorem 4. If M is (α,β)-sample accurate, then by Lemma 5, it is $(\alpha+c,\beta/c)$-distributional accurate over the posterior for $c>0$. And if M is (ε,δ)-distributional accurate (over $\mathcal{P}$, then we have:

$$\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(S)}\left[\max_{(q,a)\in\pi}\left|a-q(\mathcal{Q}_\pi)\right|>\alpha+c\right]\leq\frac{\beta}{c}\quad\text{and}\quad\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(S)}\left[\max_{(q,a)\in\pi}\left|q(\mathcal{P})-q(\mathcal{Q}_\pi)\right|>\varepsilon\right]\leq\delta.$$

Then, outside of the failure probability of at most $\beta/c+\delta$,

$$\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(S)}\left[\max_{(q,a)\in\pi}\left|a-q(\mathcal{P})\right|>\alpha+c+\varepsilon\right]\leq\frac{\beta}{c}+\delta.$$

□

3.1 Technical proof

The proof Lemma 5 will be an application of Markov's inequality along with the Bayesian resampling lemma:

Lemma 6 (Bayesian resampling lemma)**.** *Let $E \subset \mathcal{X}^n \times \mathbf{\Pi}$ be an event. Then:*

$$\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(S)}\left[(S,\pi)\in E\right] = \Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(S)\,,\,S'\sim\mathcal{Q}_\pi}\left[(S',\pi)\in E\right]$$

Proof of Lemma 5. We wish to provide an upper bound on $|a - q(\mathcal{Q}_\pi)| > \alpha'$ in probability; this suggests that we should apply a tail bound, like Markov:

$$\Pr_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(S)}\left[\max_{(q,a)\in\pi}\left|a-q(\mathcal{Q}_\pi)\right| > \alpha + c\right] \le \frac{1}{c}\cdot \mathop{\mathbb{E}}_{S\sim\mathcal{P}^n\,,\,\pi\sim\Pi(S)}\left[\max_{(q,a)\in\pi}\left|a-q(\mathcal{Q}_\pi)\right| > \alpha\right].$$

But notice that $q(\mathcal{Q}_\pi)$ is itself an expectation; it would be useful to be able to expectation over (S,π) along with the expectation over S'. We can almost do this due to the linearity of expectations, except that the absolute value is not linear. So let us split up the above event into two pieces. Then, we have:

$$\begin{aligned}\Pr_{S,\pi}\left[\max_{(q,a)\in\pi} a - q(\mathcal{Q}_\pi) > \alpha + c\right] &\le \frac{1}{c}\cdot \mathop{\mathbb{E}}_{S,\pi,S'}\left[\max\left\{\max_{(q,a)\in\pi} a - q(S') - \alpha\,,\,0\right\}\right] \\ &\le \frac{1}{c}\cdot \mathop{\mathbb{E}}_{S,\pi}\left[\Pr_{S'\sim\mathcal{Q}_\pi}\left[\max_{(q,a)\in\pi} a - q(S') - \alpha > 0\right]\right],\end{aligned}$$

where the second inequality holds because the term within the first expectation is always contained in $[0,1]$. That is because $q : \mathcal{X} \to [0,1]$. But we can now work with the probability over $S' \sim \mathcal{Q}_\pi$ using the Bayesian resampling lemma, which allows us to convert probability bound of an event with respect to the posterior to the original sample S drawn from $\mathcal{P}^n$:

$$\Pr_{S,\pi}\left[\max_{(q,a)\in\pi} a - q(S) > \alpha\right] = \Pr_{S,\pi,S'}\left[\max_{(q,a)\in\pi} a - q(S') > \alpha\right].$$

Thus, we obtain the bound:

$$\Pr_{S,\pi}\left[\max_{(q,a)\in\pi} a - q(\mathcal{Q}_\pi) > \alpha + c\right] \le \frac{1}{c}\cdot\Pr_{S,\pi}\left[\max_{(q,a)\in\pi} a - q(S) > \alpha\right].$$

The same proof holds with the event $q(\mathcal{Q}_\pi) - a > \alpha + c$. So, combining them yields the result. □

4 Transfer theorem for differential privacy

Definition 7 ((ε,δ)-differential privacy)**.** Two datasets $S, S' \in \mathcal{X}^n$ are *neighbors* if they differ in at most one coordinate. An interaction $(M,\mathcal{A})$ is (ε,δ)-*differentially private* if for all data analysts $\mathcal{A}$, pairs of neighboring datasets $S, S' \in \mathcal{X}^n$, and for all events $E \subset \mathbf{\Pi}$,

$$\Pr[\Pi(M,\mathcal{A},S)\in E] \le e^\varepsilon \cdot \Pr[\Pi(M,\mathcal{A},S')\in E] + \delta.$$

If $(M,\mathcal{A})$ is (ε,δ)-differentially private for all $\mathcal{A}$, then we say that M is (ε,δ)-differentially private.

Lemma 8 (Differential privacy implies posterior insensitivity)**.** *If $(M,\mathcal{A})$ is (ε,δ)-differentially private, then for any $\mathcal{P}$, it is is (ε',δ')-posterior insensitive for all $\varepsilon' = e^\varepsilon - 1 + 2c$ and $\delta' = \frac{\delta}{c}$ where $c > 0$.*

Corollary 9 (Transfer theorem for (ε,δ)-differential privacy)**.** *Let M be (ε,δ)-differentially private and (α,β)-sample accurate for linear queries. Then, for all analyst $\mathcal{A}$, it is (α',β')-distributionally accurate for all $\alpha' = \alpha + (e^\varepsilon - 1) + c + 2d$ and $\beta' = \frac{\beta}{c} + \frac{\delta}{d}$ for all $c, d > 0$.*

4.1 Technical proof

Proof of Lemma 8. We will proceed by contradiction, as suggested by our intuition: if M is differentially private, then knowledge of a single sample X in our dataset S should not reveal very much additional information about $\Pi(S)$. But on the other hand, if M has high posterior sensitivity, then knowing $\Pi(S)$ should reveal some information about some $X \in S$. To formalize, let $X \in_R S$ be drawn uniformly at random from our dataset S. We will compare the distributions (X, Π) and $X \otimes \Pi$, where the latter distribution is the product of the marginal distributions of (X, Π). To do so, we'll make use of the following lemma:

Lemma 10. *If M is (ε, δ)-differentially private, then for any event $E \subset \mathcal{X} \times \mathbf{\Pi}$,*

$$\Pr_{S \sim \mathcal{P}^n,\, x \sim S,\, \pi \sim \Pi(S)} \left[(x, \pi) \in E\right] \leq e^{\varepsilon} \Pr_{S \sim \mathcal{P}^n,\, x \sim \mathcal{P},\, \pi \sim \Pi(S)} \left[(x, \pi) \in E\right] + \delta$$

Suppose that M were not (α, β)-posterior insensitive. That is:

$$\Pr_{S \sim \mathcal{P}^n,\, \pi \sim \Pi(S)} \left[\max_{q \in \pi} \left|q(\mathcal{Q}_\pi) - q(\mathcal{P}^n)\right| > \alpha\right] > \beta. \tag{2}$$

Let $\mathcal{Q}_{X|\pi}$ correspond to the distribution over $\mathcal{X}$ obtained by sampling $S' \sim \mathcal{Q}_\pi$ and then sampling $X \in_R S'$ uniformly at random. Notice also that $q(\mathcal{P}^n) = q(\mathcal{P})$. Then, Equation 2 implies:

$$\Pr\left[\max_{q \in \pi} |q(\mathcal{Q}_{X|\pi}) - q(\mathcal{P})| > \alpha\right] > \beta.$$

In particular, this implies that:

$$\Pr\left[d_{\mathrm{TV}}\big(\mathcal{Q}_{X|\pi}\,,\, \mathcal{P}\big) > \alpha\right] > \beta.$$

But then, for transcripts $\pi \in \mathbf{\Pi}_\alpha$ for which this event occurs, there exists a subset $A_\pi \subset \mathcal{X}$ such that:

$$\mathcal{Q}_{X|\pi}(A_\pi) - \mathcal{P}(A_\pi) > \alpha \quad \text{or} \quad \mathcal{P}(A_\pi) - \mathcal{Q}_{X|\pi}(A_\pi) > \alpha.$$

Let $\mathbf{\Pi}^+_\alpha$ and $\mathbf{\Pi}^-_\alpha$ correspond to these events, respectively. If $\Pr\left[\pi \in \mathbf{\Pi}_\alpha\right] > \beta$, then one of $\Pr\left[\pi \in \mathbf{\Pi}^\pm_\alpha\right]$ must be at least $\beta/2$. Without loss of generality, suppose $\Pr\left[\pi \in \mathbf{\Pi}^+_\alpha\right] > \beta/2$.

Let $E = \{(x, \pi) : \pi \in \mathbf{\Pi}^+_\alpha, x \in A_\pi\}$. It then follows that:

$$\begin{aligned}\Pr_{(X,\Pi)}\left[(x,\pi) \in E\right] - \Pr_{X \otimes \Pi}\left[(x,\pi) \in E\right] &= \sum_{\pi \in \mathbf{\Pi}^+_\alpha} \Pr\left[\Pi = \pi\right] \cdot \big(\mathcal{Q}_{X|\pi}(A_\pi) - \mathcal{P}(A_\pi)\big) \\ &\geq \alpha \cdot \Pr\left[\pi \in \mathbf{\Pi}^+_\alpha\right]. \end{aligned} \tag{3}$$

On the other hand, Lemma 10 shows that:

$$\begin{aligned}\Pr_{(X,\Pi)}\left[(x,\pi) \in E\right] - \Pr_{X \otimes \Pi}\left[(x,\pi) \in E\right] &\leq (e^\varepsilon - 1) \Pr_{X \otimes \Pi}\left[(x,\pi) \in E\right] + \delta \\ &\leq (e^\varepsilon - 1) \Pr_{X \otimes \Pi}\left[\pi \in \mathbf{\Pi}^+_\alpha\right] + \delta \end{aligned} \tag{4}$$

It follows that if we let $\beta = \delta/c$ for any $c > 0$, then Equations 3 and 4 contradict when $\alpha \geq (e^\varepsilon - 1) + 2c$. Thus, $(M, \mathcal{A})$ must be $\big((e^\varepsilon - 1) + 2c, \delta/c\big)$-posterior sensitive. □