Bayesian Machine Learning for Personalization of Hearing Aid Algorithms

Bert de Vries[1,2], Alexander Ypma[1] and Tjeerd Dijkstra[1]

[1]Algorithm R&D Department
GN ReSound, Eindhoven, the Netherlands

[2]Signal Processing Systems Group
Technical University Eindhoven
www.bertdv.nl

August 2006

Problem Statement

- I have a set of HA's from Phonak, Oticon, ReSound, Widex, etc.
- Next to me I have a patient.
- Goal: Pick the best hearing aid and optimize the parameter settings for this patient.
- I am allowed to measure the patient's auditory profile and conduct listening experiments.

Problem Statement

- I have a set of HA's from Phonak, Oticon, ReSound, Widex, etc.
- Next to me I have a patient.
- Goal: Pick the best hearing aid and optimize the parameter settings for this patient.
- I am allowed to measure the patient's auditory profile and conduct listening experiments.

Problem Statement

- I have a set of HA's from Phonak, Oticon, ReSound, Widex, etc.
- Next to me I have a patient.
- Goal: Pick the best hearing aid and optimize the parameter settings for this patient.
- I am allowed to measure the patient's auditory profile and conduct listening experiments.

Problem Statement

- I have a set of HA's from Phonak, Oticon, ReSound, Widex, etc.
- Next to me I have a patient.
- Goal: Pick the best hearing aid and optimize the parameter settings for this patient.
- I am allowed to measure the patient's auditory profile and conduct listening experiments.

Outline

We will provide a solution for this problem (algorithm evaluation and fitting). Our solution features the following properties:-

- A completely general (algorithm-independent) computable solution, based on elementary mathematics.
- Complete fitting and algo evaluation results including any statistic (e.g. confidence limits etc.) in seconds.
- There is no loss of information: all patient data will be optimally processed. (more measurements $\Rightarrow$ better results).
- Lots of room for algorithm designers, auditory modelers and audiologists to be creative and affect the outcome.

Outline

We will provide a solution for this problem (algorithm evaluation and fitting). Our solution features the following properties:-

- A completely general (algorithm-independent) computable solution, based on elementary mathematics.
- Complete fitting and algo evaluation results including any statistic (e.g. confidence limits etc.) in seconds.
- There is no loss of information: all patient data will be optimally processed. (more measurements $\Rightarrow$ better results).
- Lots of room for algorithm designers, auditory modelers and audiologists to be creative and affect the outcome.

Outline

We will provide a solution for this problem (algorithm evaluation and fitting). Our solution features the following properties:-

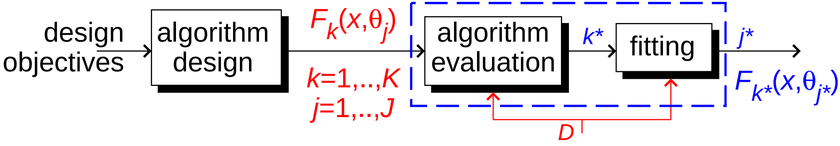
- A completely general (algorithm-independent) computable solution, based on elementary mathematics.
- Complete fitting and algo evaluation results including any statistic (e.g. confidence limits etc.) in seconds.
- There is no loss of information: all patient data will be optimally processed. (more measurements $\Rightarrow$ better results).
- Lots of room for algorithm designers, auditory modelers and audiologists to be creative and affect the outcome.

Outline

We will provide a solution for this problem (algorithm evaluation and fitting). Our solution features the following properties:-

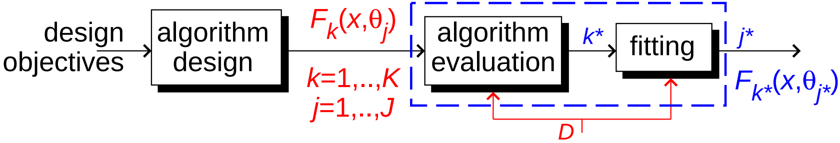
- A completely general (algorithm-independent) computable solution, based on elementary mathematics.
- Complete fitting and algo evaluation results including any statistic (e.g. confidence limits etc.) in seconds.
- There is no loss of information: all patient data will be optimally processed. (more measurements $\Rightarrow$ better results).
- Lots of room for algorithm designers, auditory modelers and audiologists to be creative and affect the outcome.

Formal Problem Statement



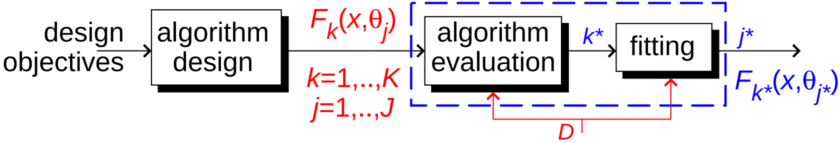
- We have K HA algorithms $\mathcal{F} = \{F_1(x,\theta),\ldots,F_K(x,\theta)\}$.
- For each algorithm F_k, there are J candidate parameter values $\Theta = \{\theta_1,\ldots,\theta_J\}$.
- We will do a set of experiments $E = \{e_1,\ldots,e_N\}$ and collect the measurements (auditory profile and listening test responses) in variable $D = \{d_1,\ldots,d_N\}$.
- Goal: Select the best algo index k^* and the best parameter value index j^*.

Formal Problem Statement



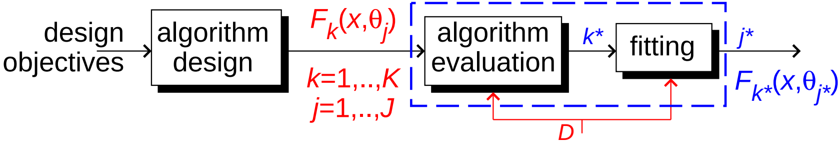
- We have K HA algorithms $\mathcal{F} = \{F_1(x,\theta), \ldots, F_K(x,\theta)\}$.
- For each algorithm F_k, there are J candidate parameter values $\Theta = \{\theta_1, \ldots, \theta_J\}$.
- We will do a set of experiments $E = \{e_1, \ldots, e_N\}$ and collect the measurements (auditory profile and listening test responses) in variable $D = \{d_1, \ldots, d_N\}$.
- Goal: Select the best algo index k^* and the best parameter value index j^*.

Formal Problem Statement



- We have K HA algorithms $\mathcal{F} = \{F_1(x,\theta),\dots,F_K(x,\theta)\}$.
- For each algorithm F_k, there are J candidate parameter values $\Theta = \{\theta_1,\dots,\theta_J\}$.
- We will do a set of experiments $E = \{e_1,\dots,e_N\}$ and collect the measurements (auditory profile and listening test responses) in variable $D = \{d_1,\dots,d_N\}$.
- Goal: Select the best algo index k^* and the best parameter value index j^*.

Formal Problem Statement



- We have K HA algorithms $\mathcal{F} = \{F_1(x,\theta), \ldots, F_K(x,\theta)\}$.
- For each algorithm F_k, there are J candidate parameter values $\Theta = \{\theta_1, \ldots, \theta_J\}$.
- We will do a set of experiments $E = \{e_1, \ldots, e_N\}$ and collect the measurements (auditory profile and listening test responses) in variable $D = \{d_1, \ldots, d_N\}$.
- Goal: Select the best algo index k^* and the best parameter value index j^*.

Rating the Candidate Parameter Values

- We cannot hope to pick the best unless we can evaluate all candidates.
- Assign a rating $\in \mathbb{R}^+$ to each candidate θ_j that expresses our belief that θ_j is the best value.
  - Think of this as an ELO rating for each candidate.
- Normalize such that the sum of the ratings equals 1.
- These ratings can now be treated as a probability distribution over the candidate value space Θ

Rating the Candidate Parameter Values

- We cannot hope to pick the best unless we can evaluate all candidates.
- Assign a rating $\in \mathbb{R}^+$ to each candidate θ_j that expresses our belief that θ_j is the best value.
  - Think of this as an ELO rating for each candidate.
- Normalize such that the sum of the ratings equals 1.
- These ratings can now be treated as a probability distribution over the candidate value space Θ

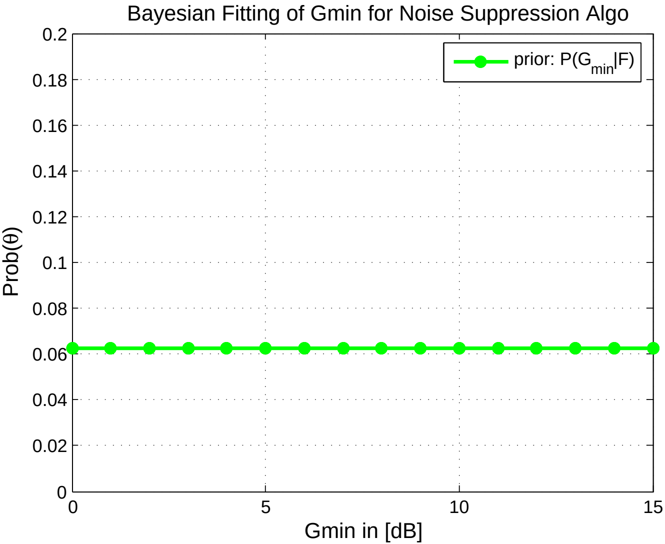
Rating the Candidate Parameter Values

- We cannot hope to pick the best unless we can evaluate all candidates.
- Assign a rating $\in \mathbb{R}^+$ to each candidate θ_j that expresses our belief that θ_j is the best value.
  - Think of this as an ELO rating for each candidate.
- Normalize such that the sum of the ratings equals 1.
- These ratings can now be treated as a probability distribution over the candidate value space Θ

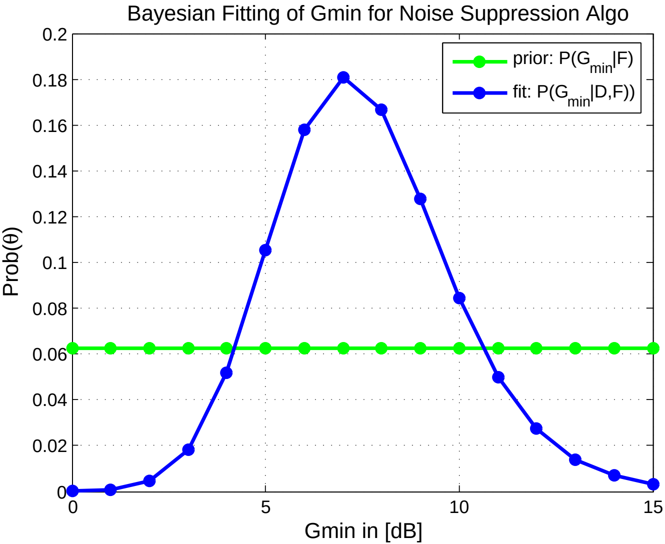
Rating the Candidate Parameter Values

- We cannot hope to pick the best unless we can evaluate all candidates.
- Assign a rating $\in \mathbb{R}^+$ to each candidate θ_j that expresses our belief that θ_j is the best value.
  - Think of this as an ELO rating for each candidate.
- Normalize such that the sum of the ratings equals 1.
- These ratings can now be treated as a probability distribution over the candidate value space Θ

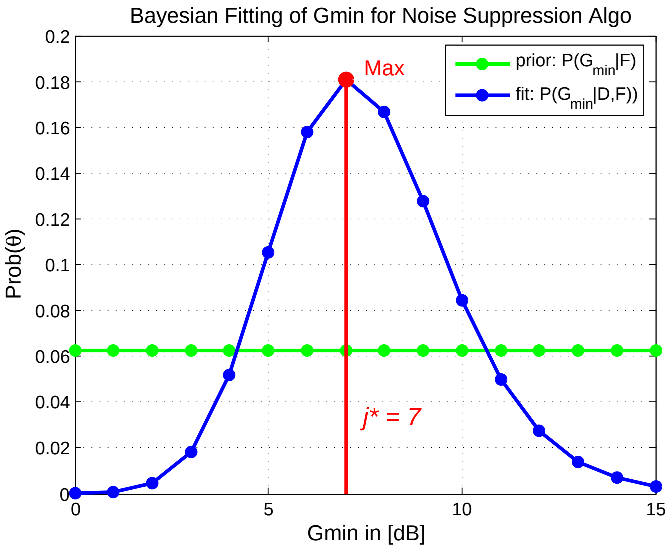
Experimental Results: Fitting G_{min} of NR Algo



Experimental Results: Fitting G_{min} of NR Algo



Experimental Results: Fitting G_{min} of NR Algo



Probability Theory

- The conditional probability $P(CR = 3|HL = 50\text{dB})$ expresses our belief that $CR = 3$, given a hearing loss of 50 (dB).
- The joint probability $P(CR = 3, HL = 50\text{dB})$ expresses our belief that $CR = 3$ and $HL = 50$ (dB).
- Sum Rule. For $\{A_1, \ldots, A_K\}$ mutually exclusive and exhaustive,

$$\boxed{\sum_{k=1}^{K} P(A_k|I) = 1}$$

- Product Rule.

$$\boxed{P(A,B|I) = P(A|B,I) \cdot P(B|I) = P(B|A,I) \cdot P(A|I)}$$

- That's all

Probability Theory

- The conditional probability $P(CR=3|HL=50\text{dB})$ expresses our belief that $CR=3$, given a hearing loss of 50 (dB).
- The joint probability $P(CR=3, HL=50\text{dB})$ expresses our belief that $CR=3$ and $HL=50$ (dB).
- Sum Rule. For $\{A_1,\ldots,A_K\}$ mutually exclusive and exhaustive,

$$\boxed{\sum_{k=1}^{K} P(A_k|I) = 1}$$

- Product Rule.

$$\boxed{P(A,B|I) = P(A|B,I)\cdot P(B|I) = P(B|A,I)\cdot P(A|I)}$$

- That's all

Probability Theory

- The conditional probability $P(CR = 3|HL = 50\text{dB})$ expresses our belief that $CR = 3$, given a hearing loss of 50 (dB).
- The joint probability $P(CR = 3, HL = 50\text{dB})$ expresses our belief that $CR = 3$ and $HL = 50$ (dB).
- Sum Rule. For $\{A_1, \ldots, A_K\}$ mutually exclusive and exhaustive,

$$\boxed{\sum_{k=1}^{K} P(A_k|I) = 1}$$

- Product Rule.

$$\boxed{P(A,B|I) = P(A|B,I) \cdot P(B|I) = P(B|A,I) \cdot P(A|I)}$$

- That's all

Probability Theory

- The conditional probability $P(CR = 3|HL = 50\text{dB})$ expresses our belief that $CR = 3$, given a hearing loss of 50 (dB).
- The joint probability $P(CR = 3, HL = 50\text{dB})$ expresses our belief that $CR = 3$ and $HL = 50$ (dB).
- Sum Rule. For $\{A_1, \ldots, A_K\}$ mutually exclusive and exhaustive,

$$\boxed{\sum_{k=1}^{K} P(A_k|I) = 1}$$

- Product Rule.

$$\boxed{P(A,B|I) = P(A|B,I) \cdot P(B|I) = P(B|A,I) \cdot P(A|I)}$$

- That's all

Probability Theory

- The conditional probability $P(CR = 3|HL = 50\text{dB})$ expresses our belief that $CR = 3$, given a hearing loss of 50 (dB).
- The joint probability $P(CR = 3, HL = 50\text{dB})$ expresses our belief that $CR = 3$ and $HL = 50$ (dB).
- Sum Rule. For $\{A_1, \dots, A_K\}$ mutually exclusive and exhaustive,

$$\boxed{\sum_{k=1}^{K} P(A_k|I) = 1}$$

- Product Rule.

$$\boxed{P(A, B|I) = P(A|B, I) \cdot P(B|I) = P(B|A, I) \cdot P(A|I)}$$

- That's all

Bayes for Fitting and Algo Evaluation

- The product rule for $P(\theta_j, D | \{E, F_k\}) =$

$$\underbrace{P(\theta_j|D,\{E,F_k\})}_{\text{fitting}} \cdot \underbrace{P(D|\{E,F_k\})}_{\textit{evaluation}} = \underbrace{P(D|\theta_j,\{E,F_k\})}_{\textit{PCM}} \cdot \underbrace{P(\theta_j|\{E,F_k\})}_{\textit{prior}}$$

- Rewrite as

$$\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$$

- We can show that

$$\begin{aligned} \mathrm{EVAL} &= P(D|\{E,F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR} \\ \mathrm{FIT} &= P(\theta_j|D,\{E,F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}} \end{aligned}$$

- End of story, in principle.

Bayes for Fitting and Algo Evaluation

- The product rule for $P(\theta_j, D|\{E, F_k\}) =$

$$\underbrace{P(\theta_j|D,\{E,F_k\})}_{\text{fitting}} \cdot \underbrace{P(D|\{E,F_k\})}_{\textit{evaluation}} = \underbrace{P(D|\theta_j,\{E,F_k\})}_{\textit{PCM}} \cdot \underbrace{P(\theta_j|\{E,F_k\})}_{\textit{prior}}$$

- Rewrite as

$$\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$$

- We can show that

$$\begin{aligned} \mathrm{EVAL} &= P(D|\{E,F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR} \\ \mathrm{FIT} &= P(\theta_j|D,\{E,F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}} \end{aligned}$$

- End of story, in principle.

Bayes for Fitting and Algo Evaluation

- The product rule for $P(\theta_j, D|\{E, F_k\}) =$

$$\underbrace{P(\theta_j|D,\{E,F_k\})}_{\text{fitting}} \cdot \underbrace{P(D|\{E,F_k\})}_{\textit{evaluation}} = \underbrace{P(D|\theta_j,\{E,F_k\})}_{\textit{PCM}} \cdot \underbrace{P(\theta_j|\{E,F_k\})}_{\textit{prior}}$$

- Rewrite as

$$\text{FIT} \times \text{EVAL} = \text{PCM} \times \text{PRIOR}$$

- We can show that

$$\begin{aligned} \text{EVAL} &= P(D|\{E,F_k\}) = \sum_j \text{PCM} \times \text{PRIOR} \\ \text{FIT} &= P(\theta_j|D,\{E,F_k\}) = \frac{\text{PCM} \times \text{PRIOR}}{\text{EVAL}} \end{aligned}$$

- End of story, in principle.

Bayes for Fitting and Algo Evaluation

- The product rule for $P(\theta_j, D | \{E, F_k\}) =$

$$\underbrace{P(\theta_j|D,\{E,F_k\})}_{\text{fitting}} \cdot \underbrace{P(D|\{E,F_k\})}_{\textit{evaluation}} = \underbrace{P(D|\theta_j,\{E,F_k\})}_{\textit{PCM}} \cdot \underbrace{P(\theta_j|\{E,F_k\})}_{\textit{prior}}$$

- Rewrite as

$$\text{FIT} \times \text{EVAL} = \text{PCM} \times \text{PRIOR}$$

- We can show that

$$\begin{aligned} \text{EVAL} &= P(D|\{E,F_k\}) = \sum_j \text{PCM} \times \text{PRIOR} \\ \text{FIT} &= P(\theta_j|D,\{E,F_k\}) = \frac{\text{PCM} \times \text{PRIOR}}{\text{EVAL}} \end{aligned}$$

- End of story, in principle.

Proof for Bayesian Fitting (OPTIONAL)

- Recall $\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$ or

$$P(\theta_j|D,\{E,F_k\}) \cdot P(D|\{E,F_k\}) = \mathrm{PCM} \times \mathrm{PRIOR}$$

- Sum over all parameter value candidates

$$\sum_j P(\theta_j|D,\{E,F_k\}) \cdot P(D|\{E,F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR}$$

- Since $\sum_j P(\theta_j|D,\{E,F_k\}) = 1$ (sum rule), it follows that

$$\mathrm{EVAL} = P(D|\{E,F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR}$$

$$\mathrm{FIT} = P(\theta_j|D,\{E,F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}}$$

- Q.E.D.

Proof for Bayesian Fitting (OPTIONAL)

- Recall $\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$ or

$$P(\theta_j|D,\{E,F_k\}) \cdot P(D|\{E,F_k\}) = \mathrm{PCM} \times \mathrm{PRIOR}$$

- Sum over all parameter value candidates

$$\sum_j P(\theta_j|D,\{E,F_k\}) \cdot P(D|\{E,F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR}$$

- Since $\sum_j P(\theta_j|D,\{E,F_k\}) = 1$ (sum rule), it follows that

$$\mathrm{EVAL} = P(D|\{E,F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR}$$

$$\mathrm{FIT} = P(\theta_j|D,\{E,F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}}$$

- Q.E.D.

Proof for Bayesian Fitting (OPTIONAL)

- Recall $\text{FIT} \times \text{EVAL} = \text{PCM} \times \text{PRIOR}$ or

$$P(\theta_j | D, \{E, F_k\}) \cdot P(D | \{E, F_k\}) = \text{PCM} \times \text{PRIOR}$$

- Sum over all parameter value candidates

$$\sum_j P(\theta_j | D, \{E, F_k\}) \cdot P(D | \{E, F_k\}) = \sum_j \text{PCM} \times \text{PRIOR}$$

- Since $\sum_j P(\theta_j | D, \{E, F_k\}) = 1$ (sum rule), it follows that

$$\text{EVAL} = P(D | \{E, F_k\}) = \sum_j \text{PCM} \times \text{PRIOR}$$

$$\text{FIT} = P(\theta_j | D, \{E, F_k\}) = \frac{\text{PCM} \times \text{PRIOR}}{\text{EVAL}}$$

- Q.E.D.

Proof for Bayesian Fitting (OPTIONAL)

- Recall $\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$ or

$$P(\theta_j|D,\{E,F_k\}) \cdot P(D|\{E,F_k\}) = \mathrm{PCM} \times \mathrm{PRIOR}$$

- Sum over all parameter value candidates

$$\sum_j P(\theta_j|D,\{E,F_k\}) \cdot P(D|\{E,F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR}$$

- Since $\sum_j P(\theta_j|D,\{E,F_k\}) = 1$ (sum rule), it follows that

$$\mathrm{EVAL} = P(D|\{E,F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR}$$

$$\mathrm{FIT} = P(\theta_j|D,\{E,F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}}$$

- Q.E.D.

The Algorithm Evaluation factor $\mathrm{EVAL} \equiv P(D|\{E, F_k\})$ (OPTIONAL)

- Technically, $P(F_k|D)$ is more accurate than $\mathrm{EVAL} \equiv P(D|F_k)$.
- Consider 2 algorithm candidates, F_k and F_l; we can evaluate F_k versus F_l by the ratio

$$\frac{P(F_k|D)}{P(F_l|D)} = \frac{\frac{P(D|F_k)P(F_k)}{P(D)}}{\frac{P(D|F_l)P(F_l)}{P(D)}} = \frac{P(D|F_k)}{P(D|F_l)} \times \frac{P(F_k)}{P(F_l)}$$
$$= \frac{\mathrm{EVAL}_k}{\mathrm{EVAL}_l}$$

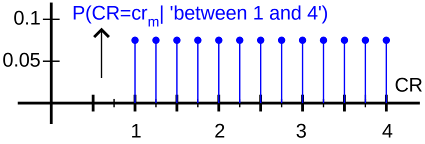
  if we have no prior preference for either F_k or F_l (i.e. $P(F_k) = P(F_l) = 0.5$).
- Hence, EVAL is a proper measure for algorithm evaluation.

More about $\mathrm{PRIOR} \equiv P(\theta_j|\{E, F_k\})$

- $P(\theta_j|\{E, F_k\})$ describes what we know (and don't know) about the parameter value candidates after design of algorithm F_k, but still before any patient measurements.
- Example: 'CR lies somewhere between 1 and 4'
- This is the place for algo engineers to shine.
- Moral: algo engineer must properly define the candidate parameter value space.
- $\mathrm{FIT} = P(\theta_j|D, \{E, F_k\})$ is like PRIOR but now after observing D.

More about $\mathrm{PRIOR} \equiv P(\theta_j | \{E, F_k\})$

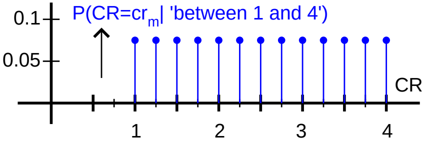
- $P(\theta_j | \{E, F_k\})$ describes what we know (and don't know) about the parameter value candidates after design of algorithm F_k, but still before any patient measurements.
- Example: 'CR lies somewhere between 1 and 4'



- This is the place for algo engineers to shine.
- Moral: algo engineer must properly define the candidate parameter value space.
- $\mathrm{FIT} = P(\theta_j | D, \{E, F_k\})$ is like PRIOR but now after observing D.

More about PRIOR $\equiv P(\theta_j|\{E, F_k\})$

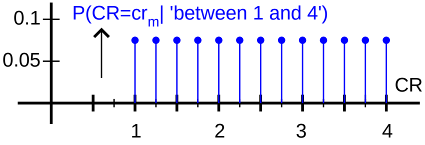
- $P(\theta_j|\{E, F_k\})$ describes what we know (and don't know) about the parameter value candidates after design of algorithm F_k, but still before any patient measurements.
- Example: 'CR lies somewhere between 1 and 4'



- This is the place for algo engineers to shine.
- Moral: algo engineer must properly define the candidate parameter value space.
- $\mathrm{FIT} = P(\theta_j|D, \{E, F_k\})$ is like PRIOR but now after observing D.

More about $\text{PRIOR} \equiv P(\theta_j|\{E, F_k\})$

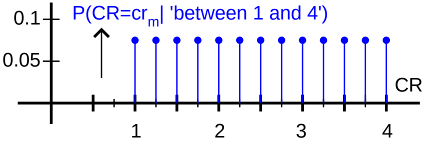
- $P(\theta_j|\{E, F_k\})$ describes what we know (and don't know) about the parameter value candidates after design of algorithm F_k, but still before any patient measurements.
- Example: 'CR lies somewhere between 1 and 4'



- This is the place for algo engineers to shine.
- Moral: algo engineer must properly define the candidate parameter value space.
- $\text{FIT} = P(\theta_j|D, \{E, F_k\})$ is like PRIOR but now after observing D.

More about $\mathrm{PRIOR} \equiv P(\theta_j|\{E, F_k\})$

- $P(\theta_j|\{E, F_k\})$ describes what we know (and don't know) about the parameter value candidates after design of algorithm F_k, but still before any patient measurements.
- Example: 'CR lies somewhere between 1 and 4'



- This is the place for algo engineers to shine.
- Moral: algo engineer must properly define the candidate parameter value space.
- $\mathrm{FIT} = P(\theta_j|D, \{E, F_k\})$ is like PRIOR but now after observing D.

Probabilistic Choice Model $\mathrm{PCM} \equiv P(D|\theta_j, \{E, F_k\})$

- $P(D|\theta_j, \{E, F_k\})$ describes what we know about patient decisions (D), given experimental design E and a given algorithm $F_k(\cdot, \theta_j)$.
- Example: Pairwise listening protocol for noise suppression.

- Bradley-Terry model

$$P(d|\theta, e, F) = \frac{e^{U(x,\theta^1)}}{e^{U(x,\theta^1)} + e^{U(x,\theta^2)}}$$

where the utility function $U(\cdot)$ is $\mathrm{PESQ}(\cdot)$, $\mathrm{PEMO}(\cdot)$, $U_{\mathrm{camfit}(\cdot)}$, $\mathrm{CSII}_3(\cdot)$ or any other utility model that you desire.

- This is the place for audiologists and auditory modelers to shine.

Probabilistic Choice Model $\mathrm{PCM} \equiv P(D|\theta_j, \{E, F_k\})$

- $P(D|\theta_j, \{E, F_k\})$ describes what we know about patient decisions (D), given experimental design E and a given algorithm $F_k(\cdot, \theta_j)$.
- Example: Pairwise listening protocol for noise suppression.

- Bradley-Terry model

$$P(d|\theta, e, F) = \frac{e^{U(x,\theta^1)}}{e^{U(x,\theta^1)} + e^{U(x,\theta^2)}}$$

where the utility function $U(\cdot)$ is $\mathrm{PESQ}(\cdot)$, $\mathrm{PEMO}(\cdot)$, $U_{\text{camfit}(\cdot)}$, $\mathrm{CSII}_3(\cdot)$ or any other utility model that you desire.

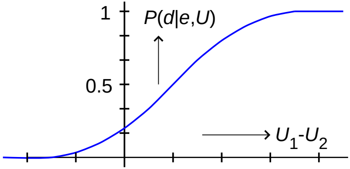
- This is the place for audiologists and auditory modelers to shine.

Probabilistic Choice Model $\mathrm{PCM} \equiv P(D|\theta_j, \{E, F_k\})$

- $P(D|\theta_j, \{E, F_k\})$ describes what we know about patient decisions (D), given experimental design E and a given algorithm $F_k(\cdot, \theta_j)$.
- Example: Pairwise listening protocol for noise suppression.

- Bradley-Terry model

$$P(d|\theta, e, F) = \frac{e^{U(x,\theta^1)}}{e^{U(x,\theta^1)} + e^{U(x,\theta^2)}}$$



where the utility function $U(\cdot)$ is $\mathrm{PESQ}(\cdot)$, $\mathrm{PEMO}(\cdot)$, $U_{\text{camfit}(\cdot)}$, $\mathrm{CSII}_3(\cdot)$ or any other utility model that you desire.

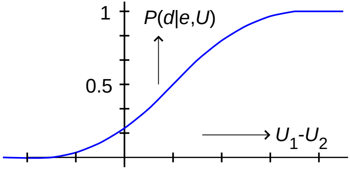
- This is the place for audiologists and auditory modelers to shine.

Probabilistic Choice Model $\mathrm{PCM} \equiv P(D|\theta_j, \{E, F_k\})$

- $P(D|\theta_j, \{E, F_k\})$ describes what we know about patient decisions (D), given experimental design E and a given algorithm $F_k(\cdot, \theta_j)$.
- Example: Pairwise listening protocol for noise suppression.

- Bradley-Terry model

$$P(d|\theta, e, F) = \frac{e^{U(x,\theta^1)}}{e^{U(x,\theta^1)} + e^{U(x,\theta^2)}}$$



where the utility function $U(\cdot)$ is $\mathrm{PESQ}(\cdot)$, $\mathrm{PEMO}(\cdot)$, $U_{\text{camfit}(\cdot)}$, $\mathrm{CSII}_3(\cdot)$ or any other utility model that you desire.

- This is the place for audiologists and auditory modelers to shine.

Bayesian Fitting and Evaluation: How to do it

1. Engineer designs HA algo $y = F_k(x, \theta)$ and specifies what he doesn't know in $\mathrm{PRIOR} \equiv P(\theta_j | F_k)$.
2. The audiologist designs experiments E and collects patient measurements D. These results are encoded in $\mathrm{PCM} \equiv P(D | E, \theta_j, F_k)$.
3. PT produces fitting and algo evaluation instantly, BOOM:

$$P(D | \{E, F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR} \qquad \text{(EVAL)}$$

$$P(\theta_j | D, \{E, F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}} \qquad \text{(FIT)}$$

- End of story, in principle

4. In a practical (HA) setting, choose

$$k^* = \arg\max_k P(D | \{E, F_k\}) \quad \text{and} \quad j^* = \arg\max_j P(\theta_j | F_{k^*}, D)$$

- End of story

Bayesian Fitting and Evaluation: How to do it

1. Engineer designs HA algo $y = F_k(x, \theta)$ and specifies what he doesn't know in $\mathrm{PRIOR} \equiv P(\theta_j | F_k)$.
2. The audiologist designs experiments E and collects patient measurements D. These results are encoded in $\mathrm{PCM} \equiv P(D | E, \theta_j, F_k)$.
3. PT produces fitting and algo evaluation instantly, BOOM:

$$P(D | \{E, F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR} \qquad \text{(EVAL)}$$

$$P(\theta_j | D, \{E, F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}} \qquad \text{(FIT)}$$

- End of story, in principle

4. In a practical (HA) setting, choose

$$k^* = \arg\max_k P(D | \{E, F_k\}) \quad \text{and} \quad j^* = \arg\max_j P(\theta_j | F_{k^*}, D)$$

- End of story

Bayesian Fitting and Evaluation: How to do it

1. Engineer designs HA algo $y = F_k(x, \theta)$ and specifies what he doesn't know in $\mathrm{PRIOR} \equiv P(\theta_j | F_k)$.
2. The audiologist designs experiments E and collects patient measurements D. These results are encoded in $\mathrm{PCM} \equiv P(D|E, \theta_j, F_k)$.
3. PT produces fitting and algo evaluation instantly, BOOM:

$$P(D|\{E, F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR} \qquad \text{(EVAL)}$$

$$P(\theta_j | D, \{E, F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}} \qquad \text{(FIT)}$$

- End of story, in principle

4. In a practical (HA) setting, choose

$$k^* = \arg\max_k P(D|\{E, F_k\}) \quad \text{and} \quad j^* = \arg\max_j P(\theta_j | F_{k^*}, D)$$

- End of story

Bayesian Fitting and Evaluation: How to do it

1. Engineer designs HA algo $y = F_k(x, \theta)$ and specifies what he doesn't know in $\mathrm{PRIOR} \equiv P(\theta_j | F_k)$.
2. The audiologist designs experiments E and collects patient measurements D. These results are encoded in $\mathrm{PCM} \equiv P(D|E, \theta_j, F_k)$.
3. PT produces fitting and algo evaluation instantly, BOOM:

$$P(D|\{E, F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR} \qquad \text{(EVAL)}$$

$$P(\theta_j|D, \{E, F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}} \qquad \text{(FIT)}$$

- End of story, in principle

4. In a practical (HA) setting, choose

$$k^* = \arg\max_k P(D|\{E, F_k\}) \quad \text{and} \quad j^* = \arg\max_j P(\theta_j | F_{k^*}, D)$$

- End of story

Bayesian Fitting and Evaluation: How to do it

1. Engineer designs HA algo $y = F_k(x, \theta)$ and specifies what he doesn't know in $\mathrm{PRIOR} \equiv P(\theta_j | F_k)$.
2. The audiologist designs experiments E and collects patient measurements D. These results are encoded in $\mathrm{PCM} \equiv P(D|E, \theta_j, F_k)$.
3. PT produces fitting and algo evaluation instantly, BOOM:

$$P(D|\{E, F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR} \qquad \text{(EVAL)}$$

$$P(\theta_j|D, \{E, F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}} \qquad \text{(FIT)}$$

- End of story, in principle

4. In a practical (HA) setting, choose

$$k^* = \arg\max_k P(D|\{E, F_k\}) \quad \text{and} \quad j^* = \arg\max_j P(\theta_j|F_{k^*}, D)$$

- End of story

Bayesian Fitting and Evaluation: How to do it

1. Engineer designs HA algo $y = F_k(x, \theta)$ and specifies what he doesn't know in $\mathrm{PRIOR} \equiv P(\theta_j | F_k)$.
2. The audiologist designs experiments E and collects patient measurements D. These results are encoded in $\mathrm{PCM} \equiv P(D|E, \theta_j, F_k)$.
3. PT produces fitting and algo evaluation instantly, BOOM:

$$P(D|\{E, F_k\}) = \sum_j \mathrm{PCM} \times \mathrm{PRIOR} \qquad \text{(EVAL)}$$

$$P(\theta_j|D, \{E, F_k\}) = \frac{\mathrm{PCM} \times \mathrm{PRIOR}}{\mathrm{EVAL}} \qquad \text{(FIT)}$$

- End of story, in principle

4. In a practical (HA) setting, choose

$$k^* = \arg\max_k P(D|\{E, F_k\}) \quad \text{and} \quad j^* = \arg\max_j P(\theta_j | F_{k^*}, D)$$

- End of story

Conservation of Information

- Recall Bayes for $\mathrm{ALL} = P(\theta_j, D | \{E, F_k\}) =$

$$P(\theta_j|D, \{E, F_k\}) \cdot P(D|\{E, F_k\}) = P(D|\theta_j, \{E, F_k\}) \cdot P(\theta_j|\{E, F_k\})$$
$$\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$$

$\Longrightarrow$ Everything we know is contained in $\mathrm{PCM} \times \mathrm{PRIOR}$ and everything we know is also exactly contained in $\mathrm{FIT} \times \mathrm{EVAL}$;

- No new information was forged and nothing was lost.
- This is not optimization; this is inference.
- k^* and j^* are approximations; they throw away information.
- Keeping track of the full PMF $\mathrm{ALL} = P(\theta_j, D | \{E, F_k\})$ provides opportunities for knowledge building, experiment selection etc.

Conservation of Information

- Recall Bayes for $\mathrm{ALL} = P(\theta_j, D|\{E, F_k\}) =$

$$P(\theta_j|D, \{E, F_k\}) \cdot P(D|\{E, F_k\}) = P(D|\theta_j, \{E, F_k\}) \cdot P(\theta_j|\{E, F_k\})$$
$$\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$$

$\Longrightarrow$ Everything we know is contained in $\mathrm{PCM} \times \mathrm{PRIOR}$ and everything we know is also exactly contained in $\mathrm{FIT} \times \mathrm{EVAL}$;

- No new information was forged and nothing was lost.
- This is not optimization; this is inference.
- k^* and j^* are approximations; they throw away information.
- Keeping track of the full PMF $\mathrm{ALL} = P(\theta_j, D|\{E, F_k\})$ provides opportunities for knowledge building, experiment selection etc.

Conservation of Information

- Recall Bayes for $\mathrm{ALL} = P(\theta_j, D|\{E, F_k\}) =$

$$P(\theta_j|D,\{E,F_k\}) \cdot P(D|\{E,F_k\}) = P(D|\theta_j,\{E,F_k\}) \cdot P(\theta_j|\{E,F_k\})$$
$$\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$$

$\Longrightarrow$ Everything we know is contained in $\mathrm{PCM} \times \mathrm{PRIOR}$ and everything we know is also exactly contained in $\mathrm{FIT} \times \mathrm{EVAL}$;

- No new information was forged and nothing was lost.
- This is not optimization; this is inference.
- k^* and j^* are approximations; they throw away information.
- Keeping track of the full PMF $\mathrm{ALL} = P(\theta_j, D|\{E, F_k\})$ provides opportunities for knowledge building, experiment selection etc.

Conservation of Information

- Recall Bayes for $\mathrm{ALL} = P(\theta_j, D | \{E, F_k\}) =$

$$P(\theta_j|D, \{E, F_k\}) \cdot P(D|\{E, F_k\}) = P(D|\theta_j, \{E, F_k\}) \cdot P(\theta_j|\{E, F_k\})$$
$$\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$$

$\Longrightarrow$ Everything we know is contained in $\mathrm{PCM} \times \mathrm{PRIOR}$ and everything we know is also exactly contained in $\mathrm{FIT} \times \mathrm{EVAL}$;

- No new information was forged and nothing was lost.
- This is not optimization; this is inference.
- k^* and j^* are approximations; they throw away information.
- Keeping track of the full PMF $\mathrm{ALL} = P(\theta_j, D | \{E, F_k\})$ provides opportunities for knowledge building, experiment selection etc.

Conservation of Information

- Recall Bayes for $\mathrm{ALL} = P(\theta_j, D | \{E, F_k\}) =$

$$P(\theta_j|D, \{E, F_k\}) \cdot P(D|\{E, F_k\}) = P(D|\theta_j, \{E, F_k\}) \cdot P(\theta_j|\{E, F_k\})$$
$$\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$$

$\Longrightarrow$ Everything we know is contained in $\mathrm{PCM} \times \mathrm{PRIOR}$ and everything we know is also exactly contained in $\mathrm{FIT} \times \mathrm{EVAL}$;

- No new information was forged and nothing was lost.
- This is not optimization; this is inference.
- k^* and j^* are approximations; they throw away information.
- Keeping track of the full PMF $\mathrm{ALL} = P(\theta_j, D | \{E, F_k\})$ provides opportunities for knowledge building, experiment selection etc.

Conservation of Information

- Recall Bayes for $\mathrm{ALL} = P(\theta_j, D|\,\{E, F_k\}) =$

$$P(\theta_j|D, \{E, F_k\}) \cdot P(D|\{E, F_k\}) = P(D|\theta_j, \{E, F_k\}) \cdot P(\theta_j|\{E, F_k\})$$
$$\mathrm{FIT} \times \mathrm{EVAL} = \mathrm{PCM} \times \mathrm{PRIOR}$$

$\Longrightarrow$ Everything we know is contained in $\mathrm{PCM} \times \mathrm{PRIOR}$ and everything we know is also exactly contained in $\mathrm{FIT} \times \mathrm{EVAL}$;

- No new information was forged and nothing was lost.
- This is not optimization; this is inference.
- k^* and j^* are approximations; they throw away information.
- Keeping track of the full PMF $\mathrm{ALL} = P(\theta_j, D|\,\{E, F_k\})$ provides opportunities for knowledge building, experiment selection etc.

It's about Pooling of Information

- The algo engineer contributes F, the audiologist E, the auditory modeler U, (and the patient D).
- Pool all contributions into the same currency (probability theory).
- All pooled information gets transferred into fitting and algo evaluation results.

$\Longrightarrow$ It's not about algorithms, it's about FEUD

- Anybody can save the day ...

It's about Pooling of Information

- The algo engineer contributes F, the audiologist E, the auditory modeler U, (and the patient D).
- Pool all contributions into the same currency (probability theory).
- All pooled information gets transferred into fitting and algo evaluation results.

$\Longrightarrow$ It's not about algorithms, it's about FEUD

- Anybody can save the day ...

It's about Pooling of Information

- The algo engineer contributes F, the audiologist E, the auditory modeler U, (and the patient D).
- Pool all contributions into the same currency (probability theory).
- All pooled information gets transferred into fitting and algo evaluation results.

$\Longrightarrow$ It's not about algorithms, it's about FEUD

- Anybody can save the day ...

It's about Pooling of Information

- The algo engineer contributes F, the auditory modeler U, (and the patient D).
- Pool all contributions into the same currency (probability theory).
- All pooled information gets transferred into fitting and algo evaluation results.

$\Longrightarrow$ It's not about algorithms, it's about FEUD

- Anybody can save the day ...

It's about Pooling of Information

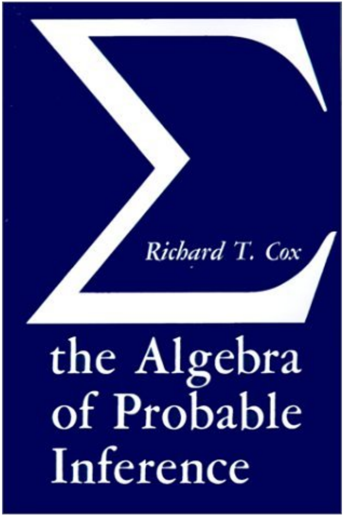
- The algo engineer contributes F, the auditory modeler U, the audiologist E, the auditory modeler U, (and the patient D).
- Pool all contributions into the same currency (probability theory).
- All pooled information gets transferred into fitting and algo evaluation results.

$\Longrightarrow$ It's not about algorithms, it's about FEUD

- Anybody can save the day ...

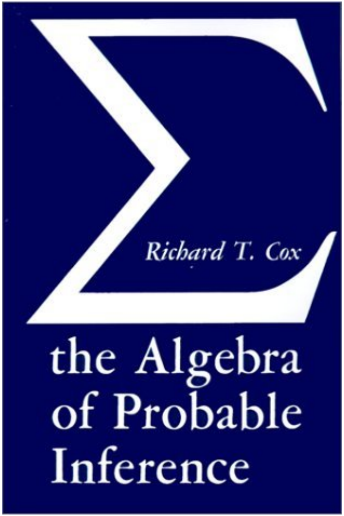
How Good is Probability Theory?

- Under some mild assumptions regarding desiderata for ‘rational reasoning’, R.T. Cox [1946] proved that the calculus of sum and product rules is both necessary and sufficient.
- (corollary). Any alternative ‘intelligent reasoning’ system is either equivalent to PT (in which case it is redundant), or inconsistent.



How Good is Probability Theory?

- Under some mild assumptions regarding desiderata for ‘rational reasoning’, R.T. Cox [1946] proved that the calculus of sum and product rules is both necessary and sufficient.
- (corollary). Any alternative ‘intelligent reasoning’ system is either equivalent to PT (in which case it is redundant), or inconsistent.



The FDA Goes Bayesian

- The FDA recently published a Draft Guidance on Bayesian statistics.
- "Bayesian statistics is a statistical theory and approach to data analysis that provides a coherent method for learning from evidence as it accumulates."

Draft Guidance
for Industry and FDA Staff

Guidance for the Use of
Bayesian Statistics in
Medical Device Clinical Trials

DRAFT GUIDANCE
This guidance document is being distributed for comment purposes only.
Draft released for comment on May 23, 2006

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Alternatively, electronic comments may be submitted to http://www.fda.gov/dockets/ecomments. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this document, contact Dr. Greg Campbell at 240-276-3133 or greg.campbell@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Biostatistics
Office of Surveillance and Biometrics

The FDA Goes Bayesian

- The FDA recently published a Draft Guidance on Bayesian statistics.
- ”Bayesian statistics is a statistical theory and approach to data analysis that provides a coherent method for learning from evidence as it accumulates.”

Draft Guidance
for Industry and FDA Staff

Guidance for the Use of
Bayesian Statistics in
Medical Device Clinical Trials

DRAFT GUIDANCE
This guidance document is being distributed for comment purposes only.
Draft released for comment on May 23, 2006

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Alternatively, electronic comments may be submitted to http://www.fda.gov/dockets/ecomments. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this document, contact Dr. Greg Campbell at 240-276-3133 or greg.campbell@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Biostatistics
Office of Surveillance and Biometrics

Conclusions

- This is the math for converting listening tests to fittings and algorithm evaluations without any loss of information.
- There is lots of room left for algorithm designers and audiologists to shine.
- No hacks; completely based on PT, which is the only consistent and coherent system for optimal processing of uncertain information.
- Thanks! For more information, visit `www.bertdv.nl`

Conclusions

- This is the math for converting listening tests to fittings and algorithm evaluations without any loss of information.
- There is lots of room left for algorithm designers and audiologists to shine.
- No hacks; completely based on PT, which is the only consistent and coherent system for optimal processing of uncertain information.
- Thanks! For more information, visit `www.bertdv.nl`

Conclusions

- This is the math for converting listening tests to fittings and algorithm evaluations without any loss of information.
- There is lots of room left for algorithm designers and audiologists to shine.
- No hacks; completely based on PT, which is the only consistent and coherent system for optimal processing of uncertain information.
- Thanks! For more information, visit `www.bertdv.nl`

Conclusions

- This is the math for converting listening tests to fittings and algorithm evaluations without any loss of information.
- There is lots of room left for algorithm designers and audiologists to shine.
- No hacks; completely based on PT, which is the only consistent and coherent system for optimal processing of uncertain information.
- Thanks! For more information, visit `www.bertdv.nl`