Imprecise Inference Models in Decision Making and Risk Analysis

Lev Utkin, St.Petersburg

September 2009, Munich

Imprecisely eight years ago

Precisely five years ago

Precisely five years ago

Precisely two months ago

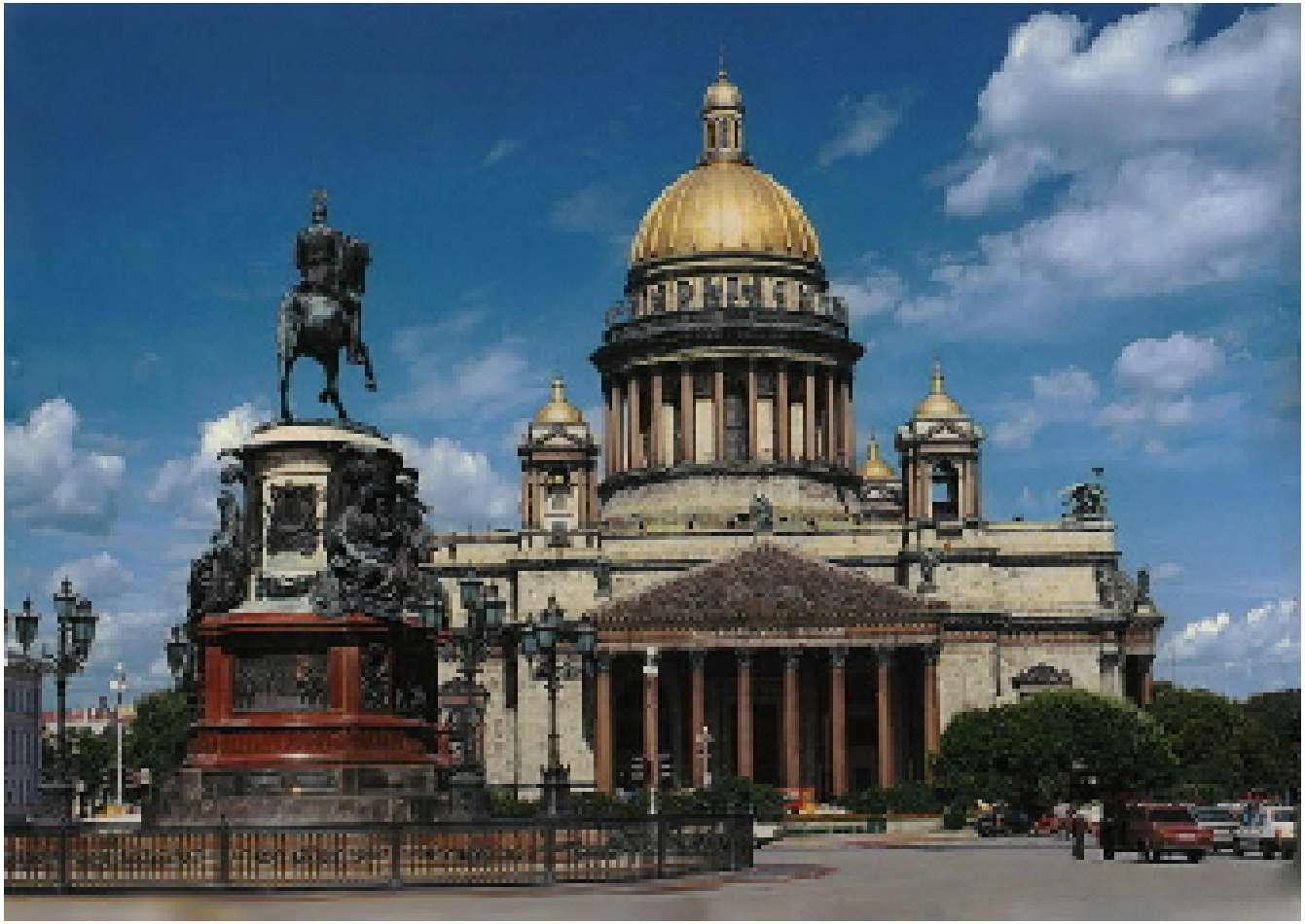
Precisely two months ago

Best greetings and happy birhday from my family

Best greetings and happy birhday from my family

Best greetings and happy birhday from Saint-Petersburg

Best greetings and happy birhday from Saint-Petersburg



Best greetings and happy birhday from Saint-Petersburg

Best greetings and happy birhday from Saint-Petersburg

Standard Bayesian Inference

1. Risk measure $\mathbb{E}X = \int_\Omega \mathbb{E}_a X \cdot \pi(a|\mathbf{b})\mathrm{d}a$.
2. **Prior distribution** $\pi(a|\mathbf{b})$ of parameters a is our opinion about the possible values a prior to collecting any information.
3. Set $\mathbf{k} = (k_1, ..., k_n)$ of observed events.
4. Likelihood function $L(\mathbf{b}|\mathbf{k}) = p(k_1|a) \cdots p(k_n|a)$.
5. **Posterior distribution** $\pi(a|\mathbf{b}, \mathbf{k}) \propto L(\mathbf{b}|\mathbf{k}) \cdot \pi(a|\mathbf{b})$ is our updated opinion about the possible values a**.**

How to choose parameters of the prior distribution?

The prior distribution is often chosen to facilitate calculation of the prior, especially through the use of *conjugate priors*. For example, the gamma distribution is a conjugate prior for the Poisson distribution.

A noninformative prior:

1. The *Bayes-Laplace postulate* or the *principle of insufficient reason* - the prior distribution should be uniform.
   - If we have no information about a, then we also have no information about $1/a$, but $1/a$ does not have a uniform distribution (not invariant under reparametrization).
   - If the parameter space is infinite, the uniform prior is improper.
   - Detailed discussion (Syversveen, 1998).
2. Other priors: Jeffreys prior (1946), Berger-Bernardo method (1989), Jaynes (1968), etc.

Basis of imprecise models

A class of the noninformative prior models is based on defining a class $\mathcal{M}$ of prior distributions π such that for event A

$$\underline{P}(A) = \inf\{P_\pi(A) : \pi \in \mathcal{M}\}, \quad \overline{P}(A) = \sup\{P_\pi(A) : \pi \in \mathcal{M}\}.$$

"Not a class of reasonable priors, but a reasonable class of priors"

- Walley's imprecise Dirichlet model (Walley 1996);
- Walley's bounded derivative model (Walley 1997);
- Imprecise models for inference in exponential families (Quaeghebeur, de Cooman 2005).

Dirichlet distribution

Assumptions (multinomial model):

1. $\Omega = \{\omega_1, ..., \omega_m\}$ is a set of possible outcomes ω_i.
2. $\Pr\{\omega_i\} = \theta_i$, $i = 1, ..., m$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)$.
3. n_i is the number of observations of ω_i in the N trials.

- The *Dirichlet* $(s, \mathbf{t})$ *prior distribution (Diri*(s, t)*)* for $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}) \propto \prod_{i=1}^{m} \theta_i^{st_i - 1}, \quad \mathbf{t} = (t_1, ..., t_m), t_i \in (0, 1).$$

- The *Dirichlet* $(N + s, \mathbf{t}^*)$ *posterior distribution (conjugate)*:

$$p(\boldsymbol{\theta}|\mathbf{n}) \propto \prod_{i=1}^{m} \theta_i^{n_i + st_i - 1}, \quad t_i^* = (n_i + st_i)/(N + s).$$

- The Beta distribution is a partial case $m = 2$.

Imprecise Dirichlet model (1)

The imprecise Dirichlet model (IDM) is (Walley 1996) the set of all Dirichlet $(s, \mathbf{t})$ *distributions such that* $\mathbf{t} \in S(1, m)$.

The *hyperparameter* s determines how quickly upper and lower probabilities of events converge as statistical data accumulate.

- Predictive probability of A

$$P(A|\mathbf{n}, \mathbf{t}, s) = \frac{n(A) + st(A)}{N + s},$$

$$n(A) = \sum_{\omega_i \in A} n_i, \quad t(A) = \sum_{\omega_i \in A} t_i.$$

Imprecise Dirichlet model (2)

- Posterior lower and upper probabilities of A:

$$\underline{P}(A|\mathbf{n}, s) = \min_{\mathbf{t} \in S(1,m)} P(A|\mathbf{n}, \mathbf{t}, s) = \frac{n(A)}{N+s},$$

$$\overline{P}(A|\mathbf{n}, s) = \max_{\mathbf{t} \in S(1,m)} P(A|\mathbf{n}, \mathbf{t}, s) = \frac{n(A)+s}{N+s}.$$

Imprecise Dirichlet model (2)

- Posterior lower and upper probabilities of A:

$$\underline{P}(A|\mathbf{n}, s) = \min_{\mathbf{t} \in S(1,m)} P(A|\mathbf{n}, \mathbf{t}, s) = \frac{n(A)}{N+s},$$

$$\overline{P}(A|\mathbf{n}, s) = \max_{\mathbf{t} \in S(1,m)} P(A|\mathbf{n}, \mathbf{t}, s) = \frac{n(A)+s}{N+s}.$$

- Before making any observations (*vacuous* model):

$$n(A) = N = 0, \underline{P}(A|\mathbf{n}, s) = 0, \overline{P}(A|\mathbf{n}, s) = 1.$$

We do not need to choose one specific prior.

Imprecise Dirichlet model (2)

- Posterior lower and upper probabilities of A:

$$\underline{P}(A|\mathbf{n}, s) = \min_{\mathbf{t} \in S(1,m)} P(A|\mathbf{n}, \mathbf{t}, s) = \frac{n(A)}{N+s},$$

$$\overline{P}(A|\mathbf{n}, s) = \max_{\mathbf{t} \in S(1,m)} P(A|\mathbf{n}, \mathbf{t}, s) = \frac{n(A)+s}{N+s}.$$

- Before making any observations (*vacuous* model):

$$n(A) = N = 0, \underline{P}(A|\mathbf{n}, s) = 0, \overline{P}(A|\mathbf{n}, s) = 1.$$

  We do not need to choose one specific prior.

- Walley 1991,1996; Bernard 2001-06, Coolen 1997, Zaffalon 2002, Hutter 2003, Quaeghebeur, de Cooman 2005

Successful applications of the IDM

1. Game-theoretic learning (Quaeghebeur, de Cooman 2003)
2. Credal treatment of missing data (Zaffalon 2002)
3. Implicative analysis for multivariate binary data (Bernard 2002)
4. Reliability analysis
   - Bayesian analysis of right-censored observations (Coolen 1997)
   - Reliability analysis of multi-state and continuum-state systems (Utkin 2006)
   - Reliability analysis of event trees (Coolen, Troffaes 2007)
   - Structural reliability analysis (Utkin, Utkin 2005)
5. Ranking procedures by pairwise comparison (Utkin 2007)
6. Analysis of NPV (Utkin 2006)
7. Decision making (Augustin, Utkin 2005)

Beta-binomial distribution

1. The number of events for time T has the binomial distribution with parameter p.
2. The conjugate distribution of p is Beta with parameters a, b.

The probability that exactly k events will occur in future by trials:

$$P(k) = \binom{N}{k} \frac{\mathrm{B}(a + k + K, b + N + N^* - k - K)}{\mathrm{B}(a + K, b + N^* - K)}.$$

Imprecise beta-binomial model

Denote $a = st$, $b = s$. Posterior lower and upper cumulative probabilities:

$$\underline{P} = \sum_{k=0}^{M} \binom{N}{k} \frac{\mathrm{B}(s+k+K, N+N^*-k-K)}{\mathrm{B}(s+K, N^*-K)},$$

$$\overline{P} = \sum_{k=0}^{M} \binom{N}{k} \frac{\mathrm{B}(k+K, s+N+N^*-k-K)}{\mathrm{B}(K, s+N^*-K)}.$$

- Before making any observations $K = N^* = 0$, $\underline{P} = 0$, $\overline{P} = 1$.
- If $s = 0$, then $\underline{P} = \overline{P}$.

Negative binomial distribution

1. The number of events for time T or τ has the Poisson distribution with parameter λ.
2. The conjugate distribution of λ is gamma with parameters a, b.

The probability that exactly k events will occur in future by time τ:

$$P(k) = \frac{\Gamma(a+K+k)}{\Gamma(a+K)k!} \cdot \left(\frac{b+T}{b+T+\tau}\right)^{a+K} \left(\frac{\tau}{b+T+\tau}\right)^{k}.$$

Imprecise negative binomial model

The set of parameters (a, b) is the triangle $(0,0)$, $(s_1, 0)$, $(0, s_2)$
-> $\mathbb{E}\lambda \in (0, \infty)$. (Coolen)
Posterior lower and upper probabilities:

$$\underline{P} = \sum_{k=0}^{M} \frac{\Gamma(s_2 + K + k)}{\Gamma(s_2 + K)k!} \cdot \left(\frac{T}{T+\tau}\right)^{s_2+K} \left(\frac{\tau}{T+\tau}\right)^{k},$$

$$\overline{P} = \sum_{k=0}^{M} \frac{\Gamma(K + k)}{\Gamma(K)k!} \cdot \left(\frac{s_1 + T}{s_1 + T + \tau}\right)^{K} \left(\frac{\tau}{s + T + \tau}\right)^{k}.$$

- Before making any observations $K = T = 0$, $\underline{P} = 0$, $\overline{P} = 1$.
- If $s_1 = s_2 = 0$, then $\underline{P} = \overline{P}$.

About two caution parameters

1. The lower bound $\underline{\mathbb{E}}_i^{(s)} X$ is

$$\underline{\mathbb{E}}_i^{(s)} X = K \frac{t}{s+T}$$

In fact, the parameter s here increases the time on the value s (hidden time).

2. The upper bound $\overline{\mathbb{E}}_i^{(s)} X$ is

$$\overline{\mathbb{E}}_i^{(s)} X = (s+K) \frac{t}{T}$$

In fact, the parameter s here increases the number of events on the value s (hidden number of events).

Gamma-exponential distribution

1. Time to an event has the exponential distribution with parameter λ.
2. The conjugate distribution of λ is gamma with parameters a, b.

The probability of an event before T:

$$P(x < T) = \left(\frac{b + T}{b + T + x} \right)^{a+n}.$$

Imprecise gamma-exponential model

The set of parameters (a, b) is the triangle $(0,0)$, $(s_1, 0)$, $(0, s_2)$
-> $\mathbb{E}\lambda \in (0, \infty)$.
Posterior lower and upper probabilities:

$$\underline{P}(x < T) = \left(\frac{T}{T+x}\right)^{s_2+n},$$

$$\overline{P}(x < T) = \left(\frac{s_1+T}{s_1+T+x}\right)^{n}.$$

- Before making any observations $T = 0$, $n = 0$, $\underline{P} = 0$, $\overline{P} = 1$.
- If $s_1 = s_2 = 0$, then $\underline{P} = \overline{P}$.

Extended IDM (ExtIDM)

- N interval-valued observations $A_i \subseteq \Omega$
- c_i is the number of occurrences of A_i, $i = 1, ..., n$.
- Belief and plausibility measures of $A \subseteq \Omega$:

$$Bel(A) = \sum_{i:A_i \subseteq A} c_i / N, \quad Pl(A) = \sum_{i:A_i \cap A \neq \varnothing} c_i / N.$$

Extended IDM (ExtIDM)

- N interval-valued observations $A_i \subseteq \Omega$
- c_i is the number of occurrences of A_i, $i = 1, ..., n$.
- Belief and plausibility measures of $A \subseteq \Omega$:

$$Bel(A) = \sum_{i:A_i \subseteq A} c_i / N, \quad Pl(A) = \sum_{i:A_i \cap A \neq \varnothing} c_i / N.$$

- Extended (ExtIDM) lower and upper probabilities:

$$\underline{P}(A|s) = \frac{N \cdot Bel(A)}{N + s}, \quad \overline{P}(A|s) = \frac{N \cdot Pl(A) + s}{N + s}.$$

Extended IDM (ExtIDM)

- N interval-valued observations $A_i \subseteq \Omega$
- c_i is the number of occurrences of A_i, $i = 1, ..., n$.
- Belief and plausibility measures of $A \subseteq \Omega$:

$$Bel(A) = \sum_{i:A_i \subseteq A} c_i / N, \quad Pl(A) = \sum_{i:A_i \cap A \neq \varnothing} c_i / N.$$

- Extended (ExtIDM) lower and upper probabilities:

$$\underline{P}(A|s) = \frac{N \cdot Bel(A)}{N+s}, \quad \overline{P}(A|s) = \frac{N \cdot Pl(A) + s}{N+s}.$$

- $\gamma = N/(N+s)$:

$$\underline{P}(A|s) = \gamma \cdot Bel(A), \quad \overline{P}(A|s) = 1 - \gamma(1 - Pl(A)).$$

Contaminated model

- ε-contaminated model is a class of probabilities which for fixed $\varepsilon \in (0,1)$ and $P(\omega_i)$ is the set $\{(1-\varepsilon)P(\omega_i) + \varepsilon Q(\omega_i)\}$ with arbitrary Q.
- Let $P(\omega_i) = c_i/N$ and $\varepsilon = s/(N+s) = 1-\gamma$.
- Then

$$\underline{P}(A|s) = (1-\varepsilon)Bel(A),$$

$$\overline{P}(A|s) = (1-\varepsilon)Pl(A) + \varepsilon.$$

Properties of the ExtIDM

1. $\underline{P}(A|s)$ and $\overline{P}(A|s)$ are belief and plausibility functions with $m^*(A_i) = c_i/(N+s)$ and $m^*(\Omega) = s/(N+s)$

Examples:

Properties of the ExtIDM

1. $\underline{P}(A|s)$ and $\overline{P}(A|s)$ are belief and plausibility functions with $m^*(A_i) = c_i/(N+s)$ and $m^*(\Omega) = s/(N+s)$
2. Total ignorance: $\underline{P}(A|s) = 0$, $\overline{P}(A|s) = 1$.

Examples:

Properties of the ExtIDM

1. $\underline{P}(A|s)$ and $\overline{P}(A|s)$ are belief and plausibility functions with $m^*(A_i) = c_i/(N+s)$ and $m^*(\Omega) = s/(N+s)$
2. Total ignorance: $\underline{P}(A|s) = 0$, $\overline{P}(A|s) = 1$.
3. $N \to \infty$: $\underline{P}(A|s) = Bel(A)$, $\overline{P}(A|s) = Pl(A)$

Examples:

Properties of the ExtIDM

1. $\underline{P}(A|s)$ and $\overline{P}(A|s)$ are belief and plausibility functions with $m^*(A_i) = c_i/(N+s)$ and $m^*(\Omega) = s/(N+s)$
2. Total ignorance: $\underline{P}(A|s) = 0$, $\overline{P}(A|s) = 1$.
3. $N \to \infty$: $\underline{P}(A|s) = Bel(A)$, $\overline{P}(A|s) = Pl(A)$

Examples:

1. One estimate $A \subset \Omega$: $Bel(A) = Pl(A) = 1$. We completely believe one estimate.

$$\underline{P}(A|1) = 1/(1+1) = 0.5, \ \overline{P}(A|1) = 1.$$

Properties of the ExtIDM

1. $\underline{P}(A|s)$ and $\overline{P}(A|s)$ are belief and plausibility functions with $m^*(A_i) = c_i/(N+s)$ and $m^*(\Omega) = s/(N+s)$
2. Total ignorance: $\underline{P}(A|s) = 0$, $\overline{P}(A|s) = 1$.
3. $N \to \infty$: $\underline{P}(A|s) = Bel(A)$, $\overline{P}(A|s) = Pl(A)$

Examples:

1. One estimate $A \subset \Omega$: $Bel(A) = Pl(A) = 1$. We completely believe one estimate.

$$\underline{P}(A|1) = 1/(1+1) = 0.5, \ \overline{P}(A|1) = 1.$$

2. N identical estimates: $Bel(A) = Pl(A) = 1$,

$$\underline{P}(A|s) = N/(N+s), \ \overline{P}(A|s) = 1.$$

Modified combination rule

Property 1 implies: extended BPAs are the **discounted** BPAs.

If $s > 0$, then $1 - K^* > 0$ in Dempster's rule, where K^* represents basic probability mass associated with conflict.

Modified Dempster's rule using extended BPAs works with conflicting evidence and conflicting evidence can always be combined by means of the rule.

Decision making and IDM (Utkin, Augustin 2005)

1. The set of actions $\mathbb{A} = \{a_1, ..., a_n\}$
2. States of nature $\Omega = \{\omega_1, ..., \omega_m\} \longleftarrow (n_1, .., n_m)$.
3. Utility function $u : (\mathbb{A} \times \Omega) \to \mathbb{R}$

1. The optimal **pure** action a_r ($\underline{\mathbb{E}}\mathbf{u}_r \geq \underline{\mathbb{E}}\mathbf{u}_k$):

$$\frac{1}{N+s}\left(\sum_{i=1}^{m} u_{ri} n_i + s \cdot \min_{i=1,...,m} u_{ri}\right) \to \max_{r=1,...,n} .$$

Decision making and IDM (Utkin, Augustin 2005)

1. The set of actions $\mathbb{A} = \{a_1, ..., a_n\}$
2. States of nature $\Omega = \{\omega_1, ..., \omega_m\} \longleftarrow (n_1, .., n_m)$.
3. Utility function $u : (\mathbb{A} \times \Omega) \to \mathbb{R}$

1. The optimal **pure** action a_r ($\underline{\mathbb{E}}\mathbf{u}_r \geq \underline{\mathbb{E}}\mathbf{u}_k$):

$$\frac{1}{N+s}\left(\sum_{i=1}^{m} u_{ri} n_i + s \cdot \min_{i=1,...,m} u_{ri}\right) \to \max_{r=1,...,n}.$$

2. The optimal **randomized** action $\lambda = (\lambda_1, .., \lambda_n)$ ($\underline{\mathbb{E}}\mathbf{u}(\lambda^*) \geq \underline{\mathbb{E}}\mathbf{u}(\lambda)$):

$$\max_{\lambda, G} G$$

subject to $G \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$,

$$G \leq \frac{1}{N+s} \sum_{r=1}^{n} \lambda_r \left(s \cdot u_{rj} + \sum_{i=1}^{m} u_{ri} \cdot n_i\right), \; j = 1, ..., m.$$

Properties of decision by IDM

$$\frac{1}{N+s}\left(\sum_{i=1}^{m} u_{ri} n_i + s \cdot \min_{i=1,\dots,m} u_{ri}\right) \to \max_{r=1,\dots,n} .$$

1. The objective function is nothing else but a mixture of two criteria:
   - the criterion of maximum expected utility with probabilities $p_i = n_i / N$ - the weight $N/(N+s)$
   - Wald's criterion - the weight $s/(N+s)$
2. When $N = 0$ (before any observation on the states of nature), Wald's criterion is used.
3. When $N \to \infty$, expected utility is used.
4. This is a frequency-based type of Hodges–Lehmann criterion.

Decision making and ExtIDM (pure action)

1. N interval-valued observations $A_i \subseteq \Omega$
2. c_i is the number of occurrences of A_i, $i = 1, ..., M$.

1. The optimal **pure** action a_r ($\underline{\mathbb{E}}\mathbf{u}_r \geq \underline{\mathbb{E}}\mathbf{u}_k$):

$$\frac{1}{N+s}\left(s \cdot \min_{i=1,\dots,m} u_{ri} + \sum_{k=1}^{M} c_k \cdot \min_{\omega_i \in A_k} u_{ri}\right) \to \max_{r=1,\dots,n}.$$

Decision making and ExtIDM (pure action)

1. N interval-valued observations $A_i \subseteq \Omega$
2. c_i is the number of occurrences of A_i, $i = 1, ..., M$.

1. The optimal **pure** action a_r ($\underline{\mathbb{E}}\mathbf{u}_r \geq \underline{\mathbb{E}}\mathbf{u}_k$):

$$\frac{1}{N+s}\left(s \cdot \min_{i=1,\ldots,m} u_{ri} + \sum_{k=1}^{M} c_k \cdot \min_{\omega_i \in A_k} u_{ri}\right) \longrightarrow \max_{r=1,\ldots,n}.$$

2. For comparison, Strat's approach (Strat 1990) ($s = 0$):

$$\left(\frac{1}{N}\sum_{k=1}^{M} c_k \cdot \min_{\omega_i \in A_k} u_{ri}\right) \longrightarrow \max_{r=1,\ldots,n}.$$

Decision making and ExtIDM (randomized action)

The optimal **randomized** action λ ($\underline{\mathbb{E}}\mathbf{u}(\lambda^*) \geq \underline{\mathbb{E}}\mathbf{u}(\lambda)$):

$$\frac{1}{N+s}\left(s \cdot V_0 + \sum_{k=1}^{M} c_k \cdot V_k\right) \to \max_{\lambda},$$

subject to $V_0, V_i \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$,

$$V_i \leq \sum_{r=1}^{n} \lambda_r u_{rj}, \;\; i = 0, ..., M, \;\; j \in J_i, \;\; J_0 = \{1, ..., m\}.$$

Advantages of decisions with IDM (1)

- Decision problem:

$$\mathbb{A} = \{a_1, a_2\}, \Omega = \{\omega_1, \omega_2\} \quad u = \begin{array}{c|c|c} & \omega_1 & \omega_2 \\ \hline a_1 & -1000 & 1 \\ \hline a_2 & 0 & 0 \end{array}$$

- There is only one judgment ($M = 1$): $A_1 = \{x_2\}$.
- Standard belief functions ($s = 0$): $a_1 \Leftarrow \underline{\mathbb{E}}\mathbf{u}_1 = 1$, $\underline{\mathbb{E}}\mathbf{u}_2 = 0$ (*too optimistic*).

Advantages of decisions with IDM (1)

- Decision problem:

$$\mathbb{A} = \{a_1, a_2\}, \Omega = \{\omega_1, \omega_2\} \quad u = \begin{array}{c|c|c} & \omega_1 & \omega_2 \\ \hline a_1 & -1000 & 1 \\ \hline a_2 & 0 & 0 \end{array}$$

- There is only one judgment ($M = 1$): $A_1 = \{x_2\}$.
- Standard belief functions ($s = 0$): $a_1 \Leftarrow \underline{\mathbb{E}}\mathbf{u}_1 = 1$, $\underline{\mathbb{E}}\mathbf{u}_2 = 0$ (*too optimistic*).
- IDM ($s = 1$): $a_2 \Leftarrow \underline{\mathbb{E}}\mathbf{u}_1 = (-1000 + 1)/2 = -499.5$, $\underline{\mathbb{E}}\mathbf{u}_2 = 0$.

Advantages of decisions with IDM (2)

- The same decision problem (but M identical observations):

$$\underline{\mathbb{E}}\mathbf{u}_1 = -\frac{s}{M+s} \cdot 1000 + \frac{M}{M+s}, \ \underline{\mathbb{E}}\mathbf{u}_2 = 0.$$

Advantages of decisions with IDM (2)

- The same decision problem (but M identical observations):

$$\underline{\mathbb{E}}\mathbf{u}_1 = -\frac{s}{M+s} \cdot 1000 + \frac{M}{M+s}, \ \underline{\mathbb{E}}\mathbf{u}_2 = 0.$$

- $s = 0$: $\underline{\mathbb{E}}\mathbf{u}_1 = 1$, $\underline{\mathbb{E}}\mathbf{u}_2 = 0$ for all $M = 1, 2, \ldots$.
  Our decision is the same irrespective of having 1 single observation or 10,000 identical observations.

Advantages of decisions with IDM (2)

- The same decision problem (but M identical observations):

$$\underline{\mathbb{E}}\mathbf{u}_1 = -\frac{s}{M+s} \cdot 1000 + \frac{M}{M+s}, \ \underline{\mathbb{E}}\mathbf{u}_2 = 0.$$

- $s = 0$: $\underline{\mathbb{E}}\mathbf{u}_1 = 1$, $\underline{\mathbb{E}}\mathbf{u}_2 = 0$ for all $M = 1, 2, \ldots$.
  Our decision is the same irrespective of having 1 single observation or 10,000 identical observations.
- $s = 1$: a_1 is only superior when $M > 1000$.

Individual risk model of insurance (problem statement)

1. N identical insurance policies for time t.
2. The claim is denoted $X_i = I_i \cdot y_i$, y_i is the claim amount.
3. Each insurance premium is c.
4. The total premium is $\Pi(t) = cN$.
5. The total amount of claims is $R(t) = X_1 + ... + X_N$.
6. The probability that aggregate claims will be less than the premium collected:

$$P = \Pr\{\Pi(t) \geq R(t)\} = \Pr\{cN \geq X_1 + ... + X_N\}.$$

Individual risk model of insurance (approaches)

If r.v. $X_1, ..., X_N$ are independent, then P is the Cdf of numbers of claims:

$$P = \sum_{k=0}^{M} p(k, w), \;\; M = \lceil \Pi(t)/y \rceil = \lceil cN/y \rceil .$$

1. $p(k, w)$ is the binomial distribution with $w = q$:

$$P = \sum_{k=0}^{M} \binom{N}{k} q^k (1-q)^{N-k}.$$

2. $p(k, w)$ is the Poisson distribution with $w = \lambda$:

$$P = \sum_{k=0}^{M} \frac{(\lambda t)^k \exp(-\lambda t)}{k!}.$$

If numbers of claims are binomially distributed

Imprecise beta-binomial model

Posterior lower and upper probabilities:

$$\underline{P} = \sum_{k=0}^{M} \binom{N}{k} \frac{\mathrm{B}(s+k+K, N+N^*-k-K)}{\mathrm{B}(s+K, N^*-K)},$$

$$\overline{P} = \sum_{k=0}^{M} \binom{N}{k} \frac{\mathrm{B}(k+K, s+N+N^*-k-K)}{\mathrm{B}(K, s+N^*-K)}.$$

If numbers of claims have the Poisson distribution

Imprecise negative binomial model

Posterior lower and upper probabilities:

$$\underline{P} = \sum_{k=0}^{M} \frac{\Gamma(s+K+k)}{\Gamma(s+K)k!} \cdot \left(\frac{T}{T+\tau}\right)^{s+K} \left(\frac{\tau}{T+\tau}\right)^{k},$$

$$\overline{P} = \sum_{k=0}^{M} \frac{\Gamma(K+k)}{\Gamma(K)k!} \cdot \left(\frac{s+T}{s+T+\tau}\right)^{K} \left(\frac{\tau}{s+T+\tau}\right)^{k}.$$

Warranty contract (problem statement)

1. n is the number of items that the buyer would like to purchase.
2. Each item is required to last for τ units of time.
3. The buyer is willing to pay x dollars per item and is prepared to tolerate at most z failures in the time interval $[0, \tau]$.
4. For each failure, the buyer needs to be compensated at the rate of y dollars per item.
5. It costs c dollars to produce a single unit of the item sold.

The *expected profit* ≥ 0 and $y = ?$

Warranty contract (approaches)

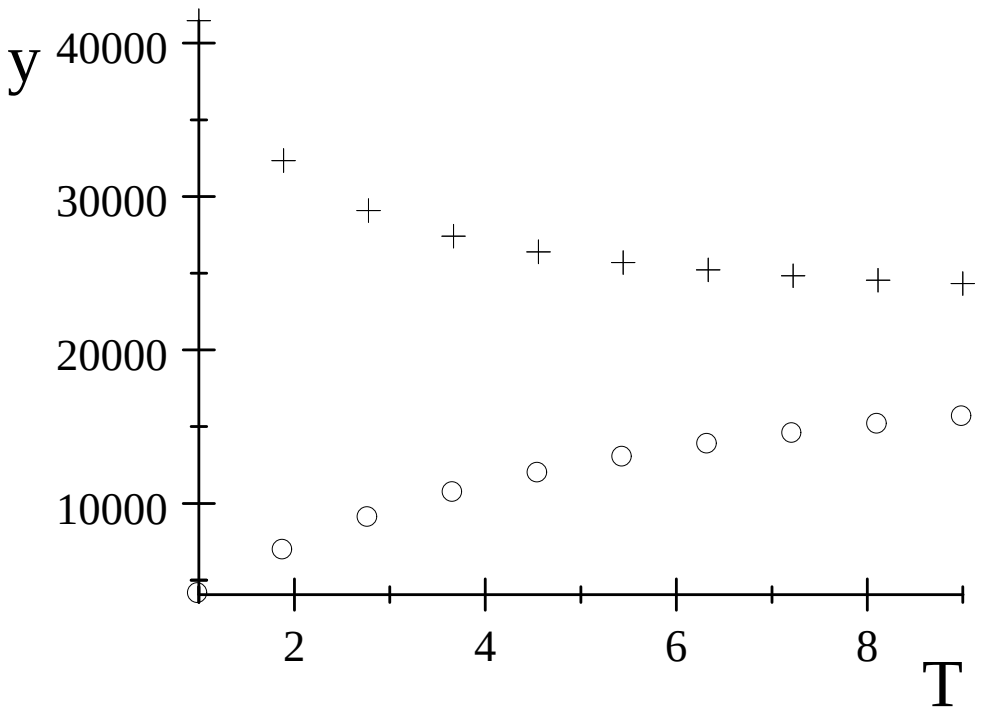
The probability of i failures in the time interval $[0, \tau]$ with the parameter λ: $p(i|\lambda)$
The *expected profit*: $\mathbb{E}_\lambda B = n(x-c) - y\sum_{i=1}^{n} i \cdot p(i|\lambda)$
By using **imprecise negative binomial model**, we have

$$\underline{\mathbb{E}}B = n(x-c) - y\sum_{k=1}^{n}\frac{\Gamma(s+K+k)}{\Gamma(s+K)(k-1)!}\left(\frac{T}{T+\tau}\right)^{s+K}\left(\frac{\tau}{T+\tau}\right)^{k}.$$

$$\overline{\mathbb{E}}B = n(x-c) - y\sum_{k=1}^{n}\frac{\Gamma(K+k)}{\Gamma(K)(k-1)!}\left(\frac{s+T}{s+T+\tau}\right)^{K}\left(\frac{\tau}{s+T+\tau}\right)^{k}.$$

Warranty contract (numerical results)

$$\underline{y} = y : \overline{\mathbb{E}} B(y) \geq 0, \quad \overline{y} = y : \underline{\mathbb{E}} B(y) \geq 0$$



Lower and upper bounds for y by $s = 1$

Problem statement

Given $\underline{F}(x) \leq F(x) \leq \overline{F}(x),\ \forall x \in \mathbb{R}$
The lower and upper expected utilities of $h(X)$:

$$\underline{\mathbb{E}}h = \inf_{\underline{F} \leq F \leq \overline{F}} \int_{\mathbb{R}} h(x)\mathrm{d}F(x),\ \overline{\mathbb{E}}h = \sup_{\underline{F} \leq F \leq \overline{F}} \int_{\mathbb{R}} h(x)\mathrm{d}F(x).$$

Approximate solution:

$$\underline{\mathbb{E}}^* h = \inf \sum_{k=1}^{N} h(x_k) z_k,\ \overline{\mathbb{E}}^* h = \sup \sum_{k=1}^{N} h(x_k) z_k,$$

subject to

$$z_k \geq 0, \quad i = 1, ..., N, \ \sum_{k=1}^{N} z_k = 1,$$

$$\sum_{k=1}^{i} z_k \leq \overline{F}(x_i),\ \sum_{k=1}^{i} z_k \geq \underline{F}(x_i),\ i = 1, ..., N.$$

Monotone utility function

1. If the function h is non-decreasing in $\mathbb{R}$, then

$$\underline{\mathbb{E}}h = \int_{\mathbb{R}} h(x)\mathrm{d}\overline{F}(x), \quad \overline{\mathbb{E}}h = \int_{\mathbb{R}} h(x)\mathrm{d}\underline{F}(x).$$

2. If the function h is non-increasing in $\mathbb{R}$, then

$$\underline{\mathbb{E}}h = \int_{\mathbb{R}} h(x)\mathrm{d}\underline{F}(x), \quad \overline{\mathbb{E}}h = \int_{\mathbb{R}} h(x)\mathrm{d}\overline{F}(x).$$

Functions having one maximum or minimum (1)

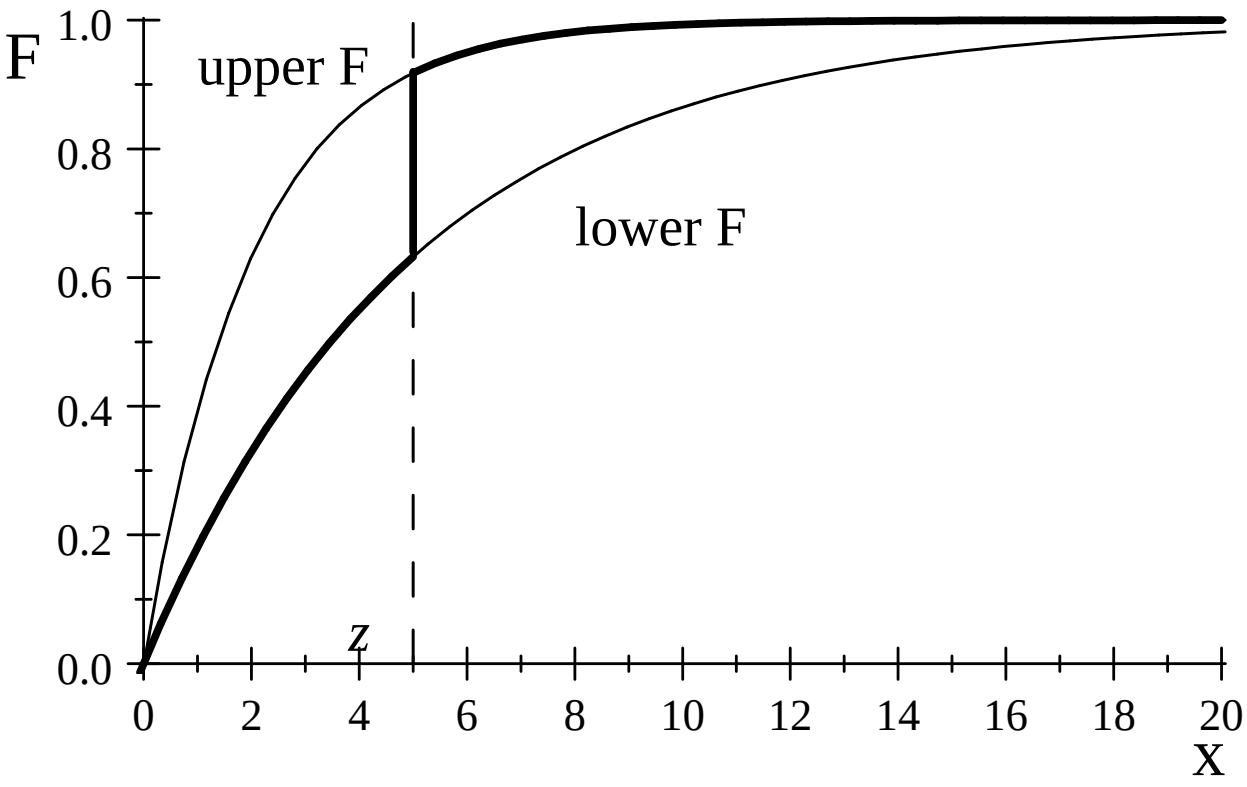
h has one maximum at point z:

$$\overline{\mathbb{E}}h = h(z)\left[\overline{F}(z) - \underline{F}(z)\right] + \int_{-\infty}^{z} h(x)\mathrm{d}\underline{F}(x) + \int_{z}^{\infty} h(x)\mathrm{d}\overline{F}(x),$$

$$\underline{\mathbb{E}}h = \min_{\alpha\in[0,1]}\left[\int_{-\infty}^{\overline{F}^{-1}(\alpha)} h(x)\mathrm{d}\overline{F}(x) + \int_{\underline{F}^{-1}(\alpha)}^{\infty} h(x)\mathrm{d}\underline{F}(x)\right],$$

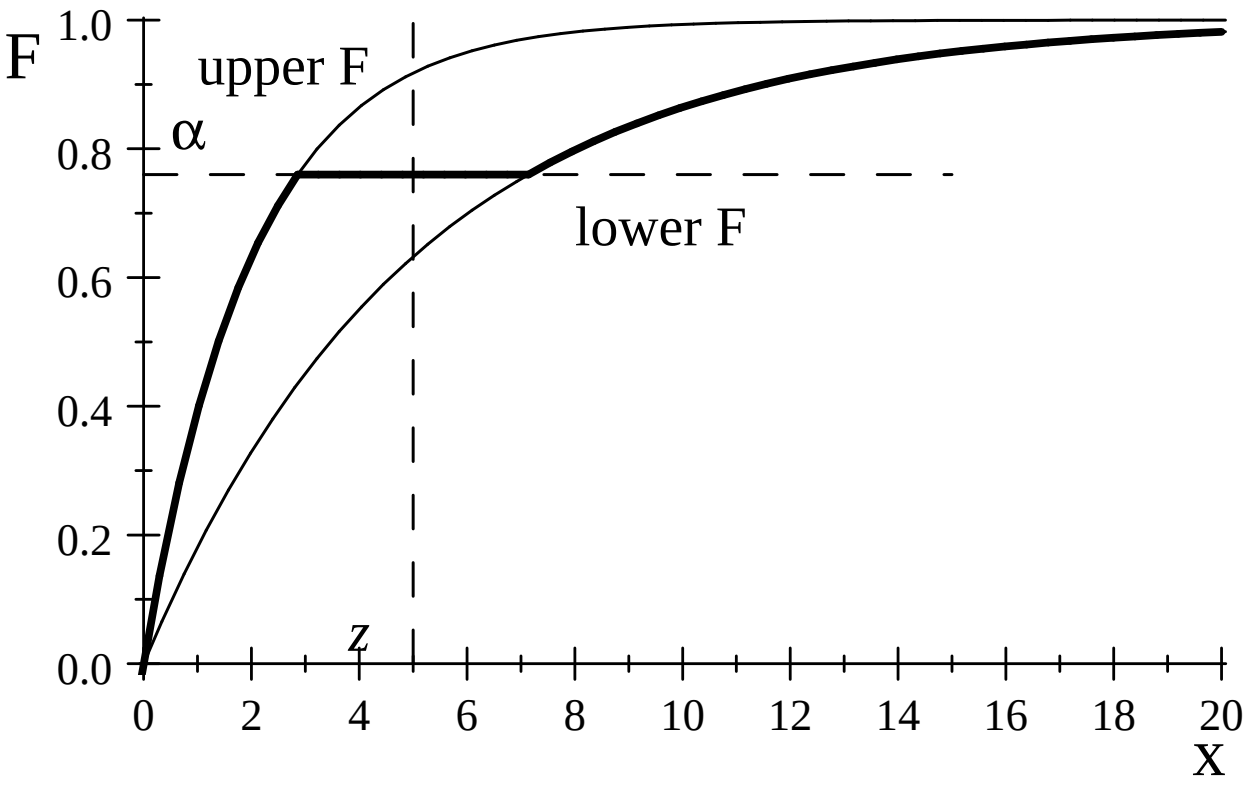
$$h\left(\overline{F}^{-1}(\alpha)\right) = h\left(\underline{F}^{-1}(\alpha)\right).$$

Functions having one maximum or minimum



The optimal distribution (thick) for computing the upper expectation

Functions having one maximum or minimum



The optimal distribution (thick) for computing the lower expectation

Possibility distributions (1)

- The possibility measure can be regarded as an upper probability measure (Dubois and Prade 1992, Walley 1996)

$$\overline{P}(A) = \Pi(A) = \sup\{\pi(x) : x \in A\}.$$

- The lower probabilities is

$$\underline{P}(A) = 1 - \Pi(A^c) = 1 - \sup\{\pi(x) : x \in A^c\}.$$

- The lower and upper probability distributions associated with the possibility distribution π:

$$\underline{F}(x) = \underline{P}((-\infty, x]) = \begin{cases} 1 - \pi(x), & x \geq x^* \\ 0, & x < x^* \end{cases},$$

$$\overline{F}(x) = \overline{P}((-\infty, x]) = \begin{cases} \pi(x), & x \leq x^* \\ 1, & x > x^* \end{cases}.$$

Possibility distributions (2)

h has one maximum:

$$\overline{\mathbb{E}}h = \begin{cases} h(z)\pi(z) + \int_z^{x^*} h(x)\mathrm{d}\pi(x), & z < x^* \\ h(z) & z = x^* \\ h(z)\pi(z) - \int_{x^*}^{z} h(x)\mathrm{d}\pi(x), & z > x^* \end{cases},$$

$$\underline{\mathbb{E}}h = \min_{\alpha\in[0,1]} \left[\int_{-\infty}^{\pi_-^{-1}(\alpha)} h(x)\mathrm{d}\pi(x) - \int_{(1-\pi_+)^{-1}(\alpha)}^{\infty} h(x)\mathrm{d}\pi(x) \right].$$

Possibility distributions of the triangular form

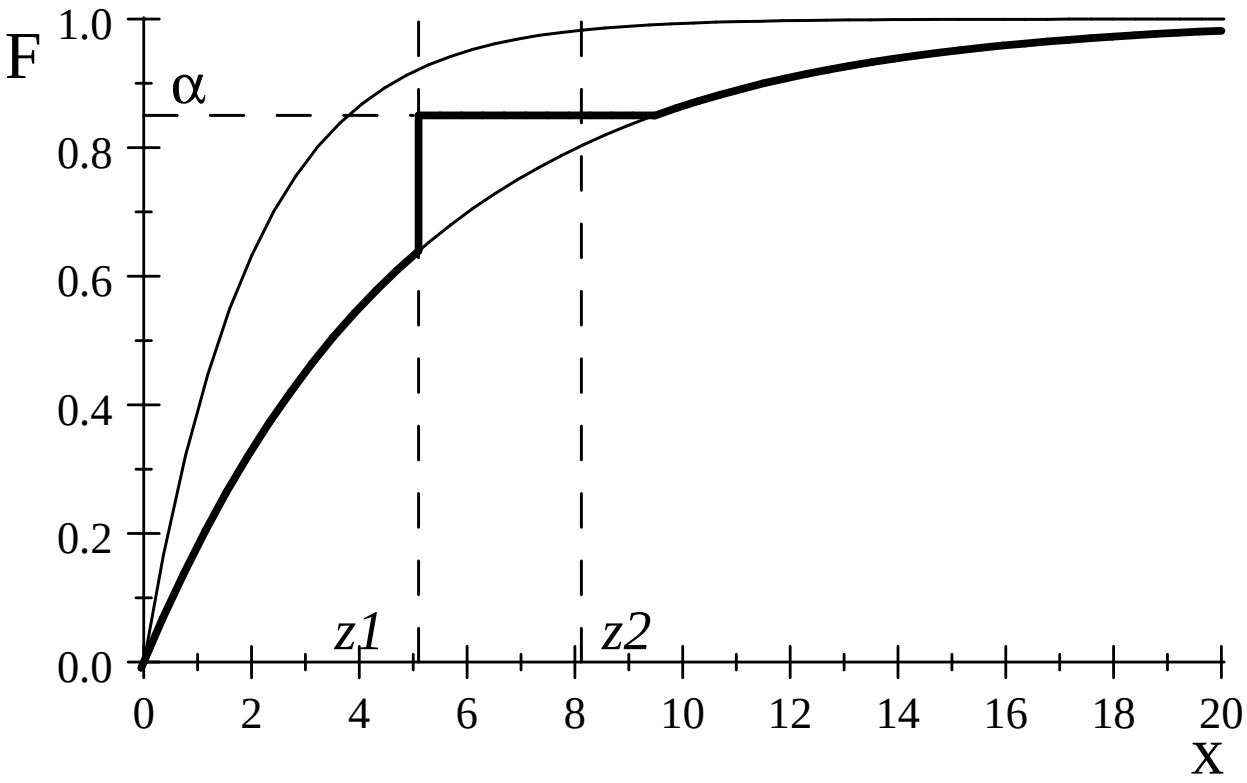
$$\pi(x) = \begin{cases} (x - a_1)/(x^* - a_1), & a_1 < x \leq x^* \\ (a_2 - x)/(a_2 - x^*), & x^* < x \leq a_2 \\ 0, & \text{otherwise} \end{cases} .$$

$$\overline{\mathbb{E}}h = \frac{1}{x^* - a_1} \left((z - a_1)h(z) + \int_z^{x^*} h(x)\mathrm{d}x \right),$$

$$\underline{\mathbb{E}}h = \frac{1}{x^* - a_1} \int_{a_1}^{\alpha(x^* - a_1) + a_1} h(x)\mathrm{d}x + \frac{1}{a_2 - x^*} \int_{\alpha(a_2 - x^*) + x^*}^{a_2} h(x)\mathrm{d}x,$$

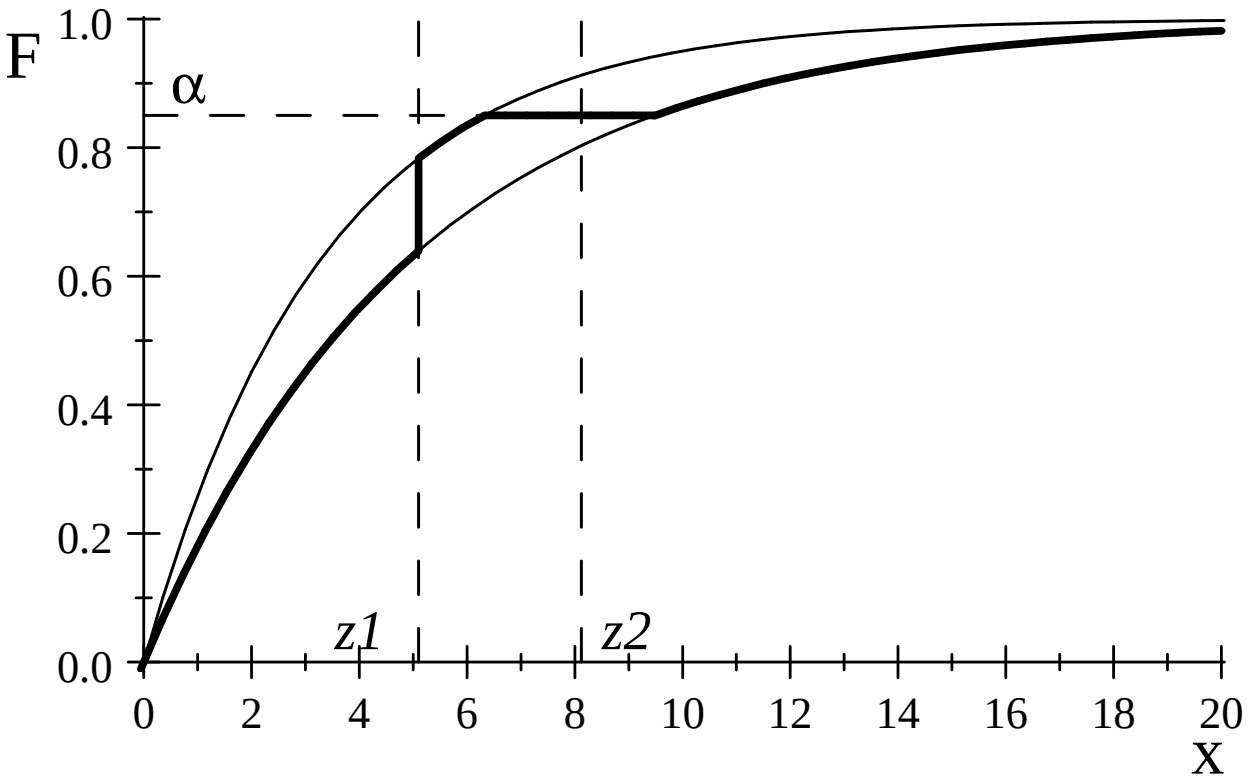
$$\alpha \Leftarrow h\left(\alpha\left(x^* - a_1\right) + a_1\right) = h\left(\alpha\left(a_2 - x^*\right) + x^*\right).$$

Functions having one maximum and one minimum



The optimal distribution (thick) for computing the upper expectation (the first case)

Functions having one maximum and one minimum



The optimal distribution (thick) for computing the upper expectation (the second case)

General case (1)

h has alternate points of the local maximum at a_i and minimum at b_{i-1}, $i = 1, 2, \ldots$.

$$b_0 < a_1 < b_1 < a_2 < b_2 < \ldots.$$

Computing $\overline{\mathbb{E}}h$ -> The optimal function $F(x)$?

1. The function $F(x) = F_i(x)$, $x \in (b_{i-1}, b_i)$, has jumps at points b_i.

General case (1)

h has alternate points of the local maximum at a_i and minimum at b_{i-1}, $i = 1, 2, \ldots$.

$$b_0 < a_1 < b_1 < a_2 < b_2 < \ldots.$$

Computing $\overline{\mathbb{E}}h$ -> The optimal function $F(x)$?

1. The function $F(x) = F_i(x)$, $x \in (b_{i-1}, b_i)$, has jumps at points b_i.
2. The size of the i-th jump is

$$\min\left(\overline{F}(b_i), \alpha_{i+1}\right) - \max\left(\underline{F}(b_i), \alpha_i\right).$$

General case (1)

h has alternate points of the local maximum at a_i and minimum at b_{i-1}, $i = 1, 2, \ldots$.

$$b_0 < a_1 < b_1 < a_2 < b_2 < \ldots.$$

Computing $\overline{\mathbb{E}}h$ -> The optimal function $F(x)$?

1. The function $F(x) = F_i(x)$, $x \in (b_{i-1}, b_i)$, has jumps at points b_i.
2. The size of the i-th jump is

$$\min\left(\overline{F}(b_i), \alpha_{i+1}\right) - \max\left(\underline{F}(b_i), \alpha_i\right).$$

3. Between jumps ...

General case (2)

Between jumps

$$F_i(x) = \begin{cases} \overline{F}(x), & x < a' \\ \alpha, & a' \leq x \leq a'' \\ \underline{F}(x), & a'' < x \end{cases},$$

where α is the root of the equation

$$h\left(\max\left(\overline{F}^{-1}(\alpha), b_{i-1}\right)\right) = h\left(\min\left(\underline{F}^{-1}(\alpha), b_i\right)\right)$$

in interval $\left[\underline{F}(a_i), \overline{F}(a_i)\right]$,

$$a' = \max\left(\overline{F}^{-1}(\alpha), b_{i-1}\right), \; a'' = \min\left(\underline{F}^{-1}(\alpha), b_i\right).$$

Questions

?