Reverse Martingales: How they Arise and how to use them for Sequential Analysis

Tudor Manole, Aaditya Ramdas

Game-Theoretic Statistical Inference, April 2021

Recall the sequential mean estimation problem

Let $(X_t)_{t=1}^{\infty}$ be a sequence of i.i.d. observations from a distribution P with mean μ. We have seen elegant approaches for sequentially testing

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0,$$

for some $\mu_0 \in \mathbb{R}$.

Recall the sequential mean estimation problem

Let $(X_t)_{t=1}^{\infty}$ be a sequence of i.i.d. observations from a distribution P with mean μ. We have seen elegant approaches for sequentially testing

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0,$$

for some $\mu_0 \in \mathbb{R}$.

- **Sequential Level-α Test** $\phi_t^{\mu_0}$: $\mathbb{P}(\exists t \geq 1 : \phi_t^{\mu_0} = 1) \leq \alpha$, under H_0.

Recall the sequential mean estimation problem

Let $(X_t)_{t=1}^{\infty}$ be a sequence of i.i.d. observations from a distribution P with mean μ. We have seen elegant approaches for sequentially testing

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0,$$

for some $\mu_0 \in \mathbb{R}$.

- **Sequential Level-α Test** $\phi_t^{\mu_0}$: $\mathbb{P}(\exists t \geq 1 : \phi_t^{\mu_0} = 1) \leq \alpha$, under H_0.
- $(1-\alpha)$**-Confidence Sequence**: $\mathbb{P}(\forall t \geq 1 : \mu \in C_t) \geq 1 - \alpha$.

Recall the sequential mean estimation problem

Let $(X_t)_{t=1}^{\infty}$ be a sequence of i.i.d. observations from a distribution P with mean μ. We have seen elegant approaches for sequentially testing

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0,$$

for some $\mu_0 \in \mathbb{R}$.

- **Sequential Level-α Test** $\phi_t^{\mu_0}$: $\mathbb{P}(\exists t \geq 1 : \phi_t^{\mu_0} = 1) \leq \alpha$, under H_0.
- $(1-\alpha)$**-Confidence Sequence**: $\mathbb{P}(\forall t \geq 1 : \mu \in C_t) \geq 1 - \alpha$.
- **Duality**: Given a sequential test $(\phi_t^{\mu_0})$ for all μ_0,

  $$C_t = \{\mu_0 : \phi_t^{\mu_0} = 0\} \text{ is a } (1-\alpha)\text{-confidence sequence,}$$

  and, given a confidence sequence (C_t),

  $$\phi_t^{\mu_0} = I(\mu_0 \notin C_t) \text{ is a level-}\alpha \text{ sequential test for } H_0.$$

A natural generalization

- **One-Sample Problem**. Given observations $(X_t)_{t=1}^{\infty}$ from a distribution P, and a known distribution Q,

$$H_0 : P = Q, \quad H_1 : P \neq Q.$$

A natural generalization

- **One-Sample Problem**. Given observations $(X_t)_{t=1}^{\infty}$ from a distribution P, and a known distribution Q,

$$H_0 : P = Q, \quad H_1 : P \neq Q.$$

Let D be a divergence, i.e. $D(P\|Q) = 0$ if $P = Q$.

A natural generalization

- **One-Sample Problem**. Given observations $(X_t)_{t=1}^{\infty}$ from a distribution P, and a known distribution Q,

$$H_0 : P = Q, \quad H_1 : P \neq Q.$$

Let D be a divergence, i.e. $D(P\|Q) = 0$ if $P = Q$. Suppose $(C_t)_{t=1}^{\infty}$ is such that

$$\mathbb{P}(\forall t \geq 1 : D(P\|Q) \in C_t) \geq 1 - \alpha.$$

Then,

$$\phi_t^Q = I(0 \notin C_t) \quad \text{is a level-}\alpha \text{ sequential test for } H_0.$$

A natural generalization

- **One-Sample Problem**. Given observations $(X_t)_{t=1}^{\infty}$ from a distribution P, and a known distribution Q,
$$H_0 : P = Q, \quad H_1 : P \neq Q.$$
Let D be a divergence, i.e. $D(P\|Q) = 0$ if $P = Q$. Suppose $(C_t)_{t=1}^{\infty}$ is such that
$$\mathbb{P}(\forall t \geq 1 : D(P\|Q) \in C_t) \geq 1 - \alpha.$$
Then,
$$\phi_t^Q = I(0 \notin C_t) \quad \text{is a level-}\alpha \text{ sequential test for } H_0.$$
- **Two-Sample Problem**. Given observations $(X_t)_{t=1}^{\infty}$ from a distribution P AND observations $(Y_s)_{s=1}^{\infty}$ from a distribution Q,
$$H_0 : P = Q, \quad H_1 : P \neq Q.$$

A natural generalization

- **One-Sample Problem**. Given observations $(X_t)_{t=1}^{\infty}$ from a distribution P, and a known distribution Q,

$$H_0 : P = Q, \quad H_1 : P \neq Q.$$

Let D be a divergence, i.e. $D(P\|Q) = 0$ if $P = Q$. Suppose $(C_t)_{t=1}^{\infty}$ is such that

$$\mathbb{P}(\forall t \geq 1 : D(P\|Q) \in C_t) \geq 1 - \alpha.$$

Then,

$$\phi_t^Q = I(0 \notin C_t) \quad \text{is a level-}\alpha \text{ sequential test for } H_0.$$

- **Two-Sample Problem**. Given observations $(X_t)_{t=1}^{\infty}$ from a distribution P AND observations $(Y_s)_{s=1}^{\infty}$ from a distribution Q,

$$H_0 : P = Q, \quad H_1 : P \neq Q.$$

Derive $(C_{ts})_{t,s=1}^{\infty}$ such that: $\mathbb{P}(\forall t, s \geq 1 : D(P\|Q) \in C_{ts}) \geq 1 - \alpha.$

Some examples of divergences

- **Distance between Means**: If P and Q are univariate,

$$D(P\|Q) = \Big|\mathbb{E}_P[X] - \mathbb{E}_Q[Y]\Big|.$$

Some examples of divergences

- **Distance between Means**: If P and Q are univariate,

$$D(P\|Q) = \Big|\mathbb{E}_P[X] - \mathbb{E}_Q[Y]\Big|.$$

Sequential inference for means is thus a special case.

Some examples of divergences

- **Distance between Means**: If P and Q are univariate,

$$D(P\|Q) = \left|\mathbb{E}_P[X] - \mathbb{E}_Q[Y]\right|.$$

  Sequential inference for means is thus a special case.

- **Distance between Multivariate Means**: More generally,

$$D(P\|Q) = \left\|\mathbb{E}_P[X] - \mathbb{E}_Q[Y]\right\|.$$

Some examples of divergences

- **Kolmogorov-Smirnov Divergence**: If P, Q are univariate, with CDFs F, G,

$$D(P\|Q) = \sup_{x\in\mathbb{R}} \left|F(x) - G(x)\right|.$$

Some examples of divergences

- **Kolmogorov-Smirnov Divergence**: If P, Q are univariate, with CDFs F, G,

$$D(P\|Q) = \sup_{x \in \mathbb{R}} \left|F(x) - G(x)\right|.$$

- **1-Wasserstein Distance**: $D(P\|Q) = \int_{\mathbb{R}} \left|F(x) - G(x)\right| dx.$

Some examples of divergences

- **Kolmogorov-Smirnov Divergence**: If P, Q are univariate, with CDFs F, G,

$$D(P\|Q) = \sup_{x \in \mathbb{R}} \left| F(x) - G(x) \right|.$$

- **1-Wasserstein Distance**: $D(P\|Q) = \int_{\mathbb{R}} \left| F(x) - G(x) \right| dx.$

- **Kernel Maximum Mean Discrepancy**: Given an RKHS $\mathcal{H}$ with kernel K,

$$D^2(P\|Q) = \|\mu_P - \mu_Q\|^2_{\mathcal{H}} = \mathbb{E}_{X,X' \sim P}[K(X,X')] + \mathbb{E}_{Y,Y' \sim Q}\left[K(Y,Y')\right] - 2\mathbb{E}_{\substack{X \sim P \\ Y \sim Q}}\left[K(X,Y)\right],$$

where μ_P and μ_Q are mean embeddings of P and Q.

Some examples of divergences

- **Kolmogorov-Smirnov Divergence**: If P, Q are univariate, with CDFs F, G,

$$D(P\|Q) = \sup_{x \in \mathbb{R}} \big|F(x) - G(x)\big|.$$

- **1-Wasserstein Distance**: $D(P\|Q) = \int_{\mathbb{R}} \big|F(x) - G(x)\big| dx.$

- **Kernel Maximum Mean Discrepancy**: Given an RKHS $\mathcal{H}$ with kernel K,

$$D^2(P\|Q) = \|\mu_P - \mu_Q\|_{\mathcal{H}}^2 = \mathbb{E}_{X,X' \sim P}[K(X,X')] + \mathbb{E}_{Y,Y' \sim Q}\big[K(Y,Y')\big] - 2\mathbb{E}_{\substack{X \sim P \\ Y \sim Q}}\big[K(X,Y)\big],$$

where μ_P and μ_Q are mean embeddings of P and Q.

These are special cases of so-called **Integral Probability Metrics**:

$$D(P\|Q) = \sup_{f \in \mathcal{F}} \big|\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f(Y)]\big|.$$

Some examples which we can sometimes handle

- **Total Variation Distance**:

$$D(P\|Q) = \sup_{B} |P(B) - Q(B)|.$$

Some examples which we can sometimes handle

- **Total Variation Distance**:

$$D(P\|Q) = \sup_{B} |P(B) - Q(B)|.$$

- **Kullback-Leibler Divergence**:

$$D(P\|Q) = \int \log\left(\frac{dP}{dQ}\right) dP.$$

Some examples which we can sometimes handle

- **Total Variation Distance**:

$$D(P\|Q) = \sup_B |P(B) - Q(B)|.$$

- **Kullback-Leibler Divergence**:

$$D(P\|Q) = \int \log\left(\frac{dP}{dQ}\right) dP.$$

- χ^2**-Divergence**:

$$D(P\|Q) = \int \left(\frac{dP}{dQ} - 1\right)^2 dQ.$$

Some examples which we can sometimes handle

- **Total Variation Distance**:

$$D(P\|Q) = \sup_B |P(B) - Q(B)|.$$

- **Kullback-Leibler Divergence**:

$$D(P\|Q) = \int \log\left(\frac{dP}{dQ}\right) dP.$$

- χ^2**-Divergence**:

$$D(P\|Q) = \int \left(\frac{dP}{dQ} - 1\right)^2 dQ.$$

These are all special cases of so-called φ**-divergences**: For a convex function φ,

$$D(P\|Q) = \int \varphi\left(\frac{dP}{dQ}\right) dQ.$$

In hindsight, the same tools can be used for certain functionals which are not divergences

Confidence sequence (C_t) for a functional Φ:

$$\mathbb{P}(\forall t \geq 1 : \Phi(P) \in C_t) \geq 1 - \alpha.$$

In hindsight, the same tools can be used for certain functionals which are not divergences

Confidence sequence (C_t) for a functional Φ:

$$\mathbb{P}(\forall t \geq 1 : \Phi(P) \in C_t) \geq 1 - \alpha.$$

- **Multilinear Functionals**: For a given nonnegative function h,

$$\Phi(P) = \mathbb{E}_{X,X' \sim P}[h(X, X')].$$

In hindsight, the same tools can be used for certain functionals which are not divergences

Confidence sequence (C_t) for a functional Φ:

$$\mathbb{P}(\forall t \geq 1 : \Phi(P) \in C_t) \geq 1 - \alpha.$$

- **Multilinear Functionals**: For a given nonnegative function h,

$$\Phi(P) = \mathbb{E}_{X,X' \sim P}[h(X, X')].$$

- **Conditional Value-at-Risk**: If P is univariate, with quantile function F^{-1},

$$\Phi(P) = \mathrm{CVaR}(P) = \frac{1}{\delta}\int_0^\delta F^{-1}(u)du = \mathbb{E}_P[X | X \leq F^{-1}(\delta)],$$

where $\delta \in (0, 1)$ and the second equality holds if P is continuous.

General approach

Given observations $(X_t)_{t=1}^{\infty}$ from P and $(Y_s)_{s=1}^{\infty}$ from Q, define the **empirical distributions**

$$P_t = \frac{1}{t}\sum_{i=1}^{t}\delta_{X_i}, \quad Q_s = \frac{1}{s}\sum_{j=1}^{s}\delta_{Y_j}.$$

General approach

Given observations $(X_t)_{t=1}^{\infty}$ from P and $(Y_s)_{s=1}^{\infty}$ from Q, define the **empirical distributions**

$$P_t = \frac{1}{t}\sum_{i=1}^{t} \delta_{X_i}, \quad Q_s = \frac{1}{s}\sum_{j=1}^{s} \delta_{Y_j}.$$

- **One-Sample Problem**. Show that $D(P_t \| Q)$ admits a martingale structure, and derive
$$\mathbb{P}(\forall t \geq 1 : -\ell_t \leq D(P\|Q) - D(P_t\|Q) \leq u_t) \geq 1 - \alpha.$$

General approach

Given observations $(X_t)_{t=1}^{\infty}$ from P and $(Y_s)_{s=1}^{\infty}$ from Q, define the **empirical distributions**

$$P_t = \frac{1}{t}\sum_{i=1}^{t} \delta_{X_i}, \quad Q_s = \frac{1}{s}\sum_{j=1}^{s} \delta_{Y_j}.$$

- **One-Sample Problem**. Show that $D(P_t \| Q)$ admits a martingale structure, and derive
$$\mathbb{P}(\forall t \geq 1 : -\ell_t \leq D(P\|Q) - D(P_t\|Q) \leq u_t) \geq 1 - \alpha.$$
- **Two-Sample Problem**. Show that $D(P_t \| Q_s)$ admits a partially ordered martingale structure, and derive
$$\mathbb{P}(\forall t, s \geq 1 : -\ell_{ts} \leq D(P\|Q) - D(P_t\|Q_s) \leq u_{ts}) \geq 1 - \alpha.$$

Outline

Forward filtrations

A **forward filtration** is an increasing sequence $(\mathcal{F}_t)_{t=0}^{\infty}$ of σ-algebras:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$$

Forward filtrations

A **forward filtration** is an increasing sequence $(\mathcal{F}_t)_{t=0}^{\infty}$ of σ-algebras:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$$

- $\mathcal{F}_t$ is intuitively the set of events that could have happened until time t.

Forward filtrations

A **forward filtration** is an increasing sequence $(\mathcal{F}_t)_{t=0}^{\infty}$ of σ-algebras:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$$

- $\mathcal{F}_t$ is intuitively the set of events that could have happened until time t.
- Think of $\mathcal{F}_t$ as the information known at time t.

Forward filtrations

A **forward filtration** is an increasing sequence $(\mathcal{F}_t)_{t=0}^{\infty}$ of σ-algebras:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$$

- $\mathcal{F}_t$ is intuitively the set of events that could have happened until time t.
- Think of $\mathcal{F}_t$ as the information known at time t.
- **Canonical filtration**:

$$\mathcal{C}_0 = \{\emptyset, \Omega\}, \quad \mathcal{C}_t = \sigma(X_1, \dots, X_t).$$

$\mathcal{C}_t$ is the smallest σ-algebra containing all events of the form $\{X_i \in B\}$, $i = 1, \dots, t$.

Forward filtrations

A **forward filtration** is an increasing sequence $(\mathcal{F}_t)_{t=0}^{\infty}$ of σ-algebras:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$$

- $\mathcal{F}_t$ is intuitively the set of events that could have happened until time t.
- Think of $\mathcal{F}_t$ as the information known at time t.
- **Canonical filtration**:

$$\mathcal{C}_0 = \{\emptyset, \Omega\}, \quad \mathcal{C}_t = \sigma(X_1, \dots, X_t).$$

  $\mathcal{C}_t$ is the smallest σ-algebra containing all events of the form $\{X_i \in B\}$, $i = 1, \dots, t$.

The conditional expectation of an RV Y given $\mathcal{F}_t$ is denoted $\mathbb{E}[Y|\mathcal{F}_t]$.

$\mathbb{E}[Y|\mathcal{F}_t]$ **is our best guess of** Y **given the information contained in** $\mathcal{F}_t$**.**

For instance, $\mathbb{E}[Y|\mathcal{C}_t] = \mathbb{E}[Y|X_1, \dots, X_t]$ in the usual sense.

Forward martingales

A **forward martingale** with respect to a filtration $(\mathcal{F}_t)_{t=0}^{\infty}$ is a process $(S_t)_{t=0}^{\infty}$ such that:

1. S_t is $\mathcal{F}_t$-measurable: "S_t is constructed based only on information in $\mathcal{F}_t$".

Forward martingales

A **forward martingale** with respect to a filtration $(\mathcal{F}_t)_{t=0}^{\infty}$ is a process $(S_t)_{t=0}^{\infty}$ such that:

1. S_t is $\mathcal{F}_t$-measurable: "S_t is constructed based only on information in $\mathcal{F}_t$".
2. We have,

$$\mathbb{E}[S_{t+1}|\mathcal{F}_t] = S_t, \quad t = 0, 1, \ldots$$

Forward martingales

A **forward martingale** with respect to a filtration $(\mathcal{F}_t)_{t=0}^{\infty}$ is a process $(S_t)_{t=0}^{\infty}$ such that:

1. S_t is $\mathcal{F}_t$-measurable: "S_t is constructed based only on information in $\mathcal{F}_t$".
2. We have,
$$\mathbb{E}[S_{t+1}|\mathcal{F}_t] = S_t, \quad t = 0, 1, \ldots$$

If $\mathcal{F}_t = \mathcal{C}_t$, this reduces to the usual definition:

1. S_t is some function of $X_1, \ldots, X_t$.
2. We have,
$$\mathbb{E}[S_{t+1}|X_1, \ldots, X_t] = S_t, \quad t = 0, 1, \ldots$$

Forward martingales

A **forward martingale** with respect to a filtration $(\mathcal{F}_t)_{t=0}^{\infty}$ is a process $(S_t)_{t=0}^{\infty}$ such that:

1. S_t is $\mathcal{F}_t$-measurable: "S_t is constructed based only on information in $\mathcal{F}_t$".
2. We have,

$$\mathbb{E}[S_{t+1}|\mathcal{F}_t] = S_t, \quad t = 0, 1, \ldots$$

Similarly,

- Forward **supermartingale**: $\mathbb{E}[S_{t+1}|\mathcal{F}_t] \leq S_t$ ("decreasing with time").
- Forward **submartingale**: $\mathbb{E}[S_{t+1}|\mathcal{F}_t] \geq S_t$ ("increasing with time").

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

- **Sums**: $S_t = \sum_{i=1}^{t} X_i$ is a martingale.

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

- **Sums**: $S_t = \sum_{i=1}^t X_i$ is a martingale.

  Conversely, if (S_t) is a mean-zero martingale w.r.t. $(\mathcal{F}_t)$, then there exist (Y_t) such that $\mathbb{E}[Y_{t+1}|\mathcal{F}_t] = 0$ and $S_t = \sum_{i=1}^t Y_i$.

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

- **Sums**: $S_t = \sum_{i=1}^t X_i$ is a martingale.

  Conversely, if (S_t) is a mean-zero martingale w.r.t. $(\mathcal{F}_t)$, then there exist (Y_t) such that $\mathbb{E}[Y_{t+1}|\mathcal{F}_t] = 0$ and $S_t = \sum_{i=1}^t Y_i$.

  Proof. Take $Y_t = S_t - S_{t-1}$.

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

- **Sums**: $S_t = \sum_{i=1}^{t} X_i$ is a martingale.
- **Nonnegative Products**: $L_t = \prod_{i=1}^{t}(1 + X_i)$ is a nonneg. martingale if $X_t \geq -1$.

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

- **Sums**: $S_t = \sum_{i=1}^t X_i$ is a martingale.
- **Nonnegative Products**: $L_t = \prod_{i=1}^t (1 + X_i)$ is a nonneg. martingale if $X_t \geq -1$.

  Conversely, if (L_t) is a mean-one martingale w.r.t. $(\mathcal{F}_t)$, then there exist $Y_t \geq -1$ such that $\mathbb{E}[Y_{t+1}|\mathcal{F}_t] = 0$ and $S_t = \prod_{i=1}^t (1 + Y_i)$.

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

- **Sums**: $S_t = \sum_{i=1}^t X_i$ is a martingale.
- **Nonnegative Products**: $L_t = \prod_{i=1}^t (1 + X_i)$ is a nonneg. martingale if $X_t \geq -1$.

  Conversely, if (L_t) is a mean-one martingale w.r.t. $(\mathcal{F}_t)$, then there exist $Y_t \geq -1$ such that $\mathbb{E}[Y_{t+1}|\mathcal{F}_t] = 0$ and $S_t = \prod_{i=1}^t (1 + Y_i)$.

  Proof. Take $$Y_t = \frac{L_t}{L_{t-1}} = 1 + \underbrace{\frac{L_t - L_{t-1}}{L_{t-1}}}_{X_t}. \quad (0/0 := 1)$$

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

- **Sums**: $S_t = \sum_{i=1}^t X_i$ is a martingale.
- **Nonnegative Products**: $L_t = \prod_{i=1}^t (1 + X_i)$ is a nonneg. martingale if $X_t \geq -1$.

  Conversely, if (L_t) is a mean-one martingale w.r.t. $(\mathcal{F}_t)$, then there exist $Y_t \geq -1$ such that $\mathbb{E}[Y_{t+1}|\mathcal{F}_t] = 0$ and $S_t = \prod_{i=1}^t (1 + Y_i)$.

  We have seen similar characterizations for:

  - **Capital Processes**: $\prod_{i=1}^t (1 + \lambda_i X_i)$ with (λ_t) predictable, i.e. λ_t is $\mathcal{F}_{t-1}$-measurable.

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

- **Sums**: $S_t = \sum_{i=1}^t X_i$ is a martingale.
- **Nonnegative Products**: $L_t = \prod_{i=1}^t (1 + X_i)$ is a nonneg. martingale if $X_t \geq -1$.

  Conversely, if (L_t) is a mean-one martingale w.r.t. $(\mathcal{F}_t)$, then there exist $Y_t \geq -1$ such that $\mathbb{E}[Y_{t+1}|\mathcal{F}_t] = 0$ and $S_t = \prod_{i=1}^t (1 + Y_i)$.

  We have seen similar characterizations for:
  - **Capital Processes**: $\prod_{i=1}^t (1 + \lambda_i X_i)$ with (λ_t) predictable, i.e. λ_t is $\mathcal{F}_{t-1}$-measurable.
  - **Likelihood Ratios**: $\prod_{i=1}^t q(X_i)/p(X_i)$.

Two key characterizations of martingales

Let (X_t) satisfy $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = 0$.

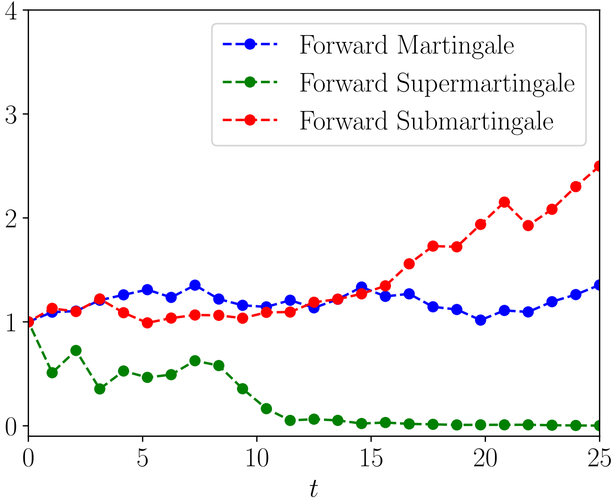
- **Sums**: $S_t = \sum_{i=1}^t X_i$ is a martingale.
- **Nonnegative Products**: $L_t = \prod_{i=1}^t (1 + X_i)$ is a nonneg. martingale if $X_t \geq -1$.

  More generally,

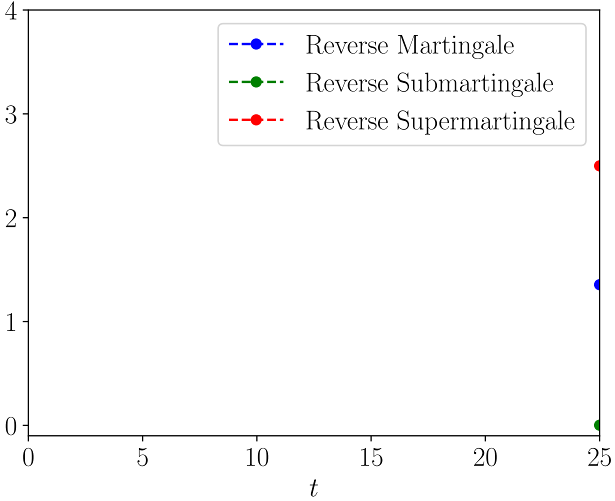
$$\mathbb{E}[X_{t+1}|\mathcal{F}_t] \leq 0 \quad \implies (L_t), (S_t) \text{ are supermartingales w.r.t. } (\mathcal{F}_t)$$
$$\mathbb{E}[X_{t+1}|\mathcal{F}_t] \geq 0 \quad \implies (L_t), (S_t) \text{ are submartingales w.r.t. } (\mathcal{F}_t).$$

Reverse Martingales

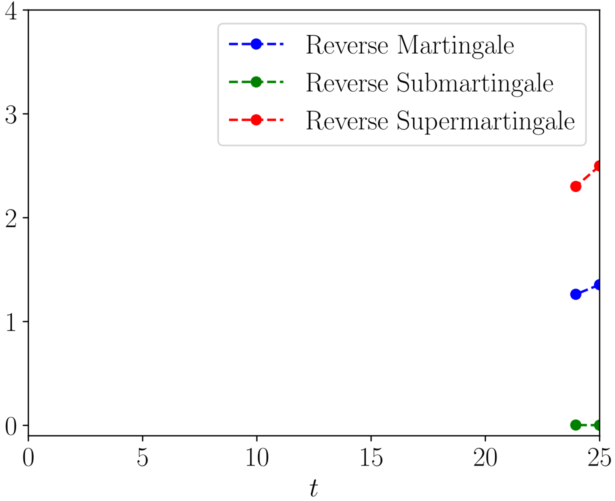
From forward to reverse martingales



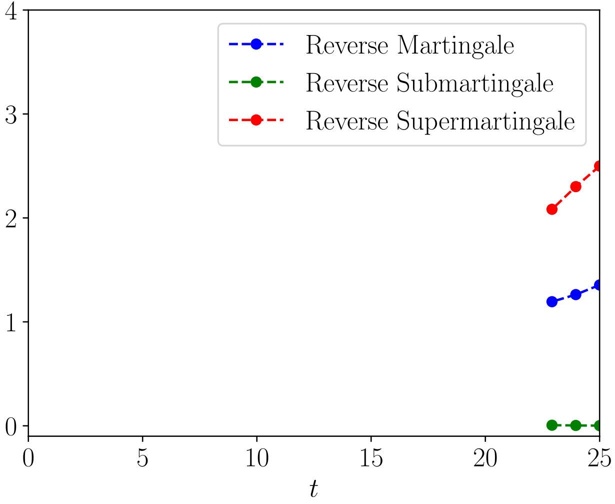
From forward to reverse martingales



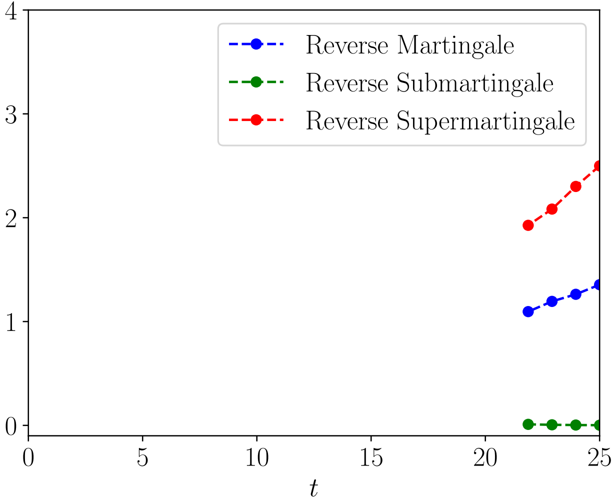
From forward to reverse martingales



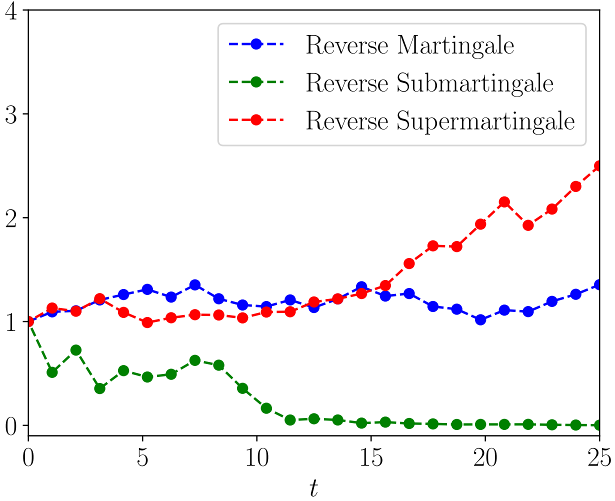
From forward to reverse martingales



From forward to reverse martingales



From forward to reverse martingales



Reverse martingales

A **reverse filtration** is a decreasing sequence $(\mathcal{R}_t)_{t=1}^{\infty}$ of σ-algebras:

$$\mathcal{R}_1 \supseteq \mathcal{R}_2 \supseteq \dots$$

Reverse martingales

A **reverse filtration** is a decreasing sequence $(\mathcal{R}_t)_{t=1}^{\infty}$ of σ-algebras:

$$\mathcal{R}_1 \supseteq \mathcal{R}_2 \supseteq \dots$$

A **reverse martingale** with respect to $(\mathcal{R}_t)_{t=1}^{\infty}$ is a process $(M_t)_{t=1}^{\infty}$ such that:

1. M_t is $\mathcal{R}_t$-measurable for all $t \geq 1$.

Reverse martingales

A **reverse filtration** is a decreasing sequence $(\mathcal{R}_t)_{t=1}^{\infty}$ of σ-algebras:

$$\mathcal{R}_1 \supseteq \mathcal{R}_2 \supseteq \dots$$

A **reverse martingale** with respect to $(\mathcal{R}_t)_{t=1}^{\infty}$ is a process $(M_t)_{t=1}^{\infty}$ such that:

1. M_t is $\mathcal{R}_t$-measurable for all $t \geq 1$.
2. We have,

$$\mathbb{E}[M_t | \mathcal{R}_{t+1}] = M_{t+1}, \quad t = 1, 2 \dots.$$

(Compare to $\mathbb{E}[S_t | \mathcal{F}_{t-1}] = S_{t-1}$ for forward martingales.)

Reverse martingales

A **reverse filtration** is a decreasing sequence $(\mathcal{R}_t)_{t=1}^{\infty}$ of σ-algebras:

$$\mathcal{R}_1 \supseteq \mathcal{R}_2 \supseteq \dots$$

A **reverse martingale** with respect to $(\mathcal{R}_t)_{t=1}^{\infty}$ is a process $(M_t)_{t=1}^{\infty}$ such that:

1. M_t is $\mathcal{R}_t$-measurable for all $t \geq 1$.
2. We have,

$$\mathbb{E}[M_t | \mathcal{R}_{t+1}] = M_{t+1}, \quad t = 1, 2 \dots .$$

   (Compare to $\mathbb{E}[S_t | \mathcal{F}_{t-1}] = S_{t-1}$ for forward martingales.)

Furthermore,

- Reverse **supermartingale**: $\mathbb{E}[M_t | \mathcal{R}_{t+1}] \leq M_{t+1}$ (compare to $\mathbb{E}[S_t | \mathcal{F}_{t-1}] \leq S_{t-1}$).
- Reverse **submartingale**: $\mathbb{E}[M_t | \mathcal{R}_{t+1}] \geq M_{t+1}$ (compare to $\mathbb{E}[S_t | \mathcal{F}_{t-1}] \geq S_{t-1}$).

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale.

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale.

Before proving this, let us try to guess the relevant reverse filtration.

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale.

Before proving this, let us try to guess the relevant reverse filtration.

Remark. If (M_t) is a reverse martingale w.r.t. $(\mathcal{G}_t)$, then it is also a reverse martingale w.r.t.

$$\mathcal{R}_t = \sigma(M_t, M_{t+1}, \dots), \quad t = 1, 2, \dots$$

Proof. $(\mathcal{R}_t)$ is the smallest filtration to which (M_t) is adapted, i.e. $\mathcal{R}_t \subseteq \mathcal{G}_t$. Thus,

$$\mathbb{E}[M_t|\mathcal{R}_{t+1}] = \mathbb{E}\big[\mathbb{E}[M_t|\mathcal{G}_{t+1}]\big|\mathcal{R}_{t+1}\big] = \mathbb{E}[M_{t+1}|\mathcal{R}_{t+1}] = M_{t+1}.$$

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale.

Before proving this, let us try to guess the relevant reverse filtration.

Remark. If (M_t) is a reverse martingale w.r.t. $(\mathcal{G}_t)$, then it is also a reverse martingale w.r.t.

$$\mathcal{R}_t = \sigma(M_t, M_{t+1}, \dots), \quad t = 1, 2, \dots$$

What is $\mathcal{R}_t$? If we know M_t and M_{t+1}, then we also know:

$$X_{t+1} = (t+1)M_{t+1} - tM_t = \sum_{i=1}^{t+1} X_i - \sum_{i=1}^{t} X_i = X_{t+1}.$$

Thus:

$$\mathcal{R}_t = \sigma(M_t, X_{t+1}, X_{t+2}, \dots).$$

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale with respect to the filtration:

$$\mathcal{R}_t = \sigma(M_t, X_{t+1}, X_{t+2}, \dots), \quad t = 1, 2, \dots$$

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale with respect to the filtration:

$$\mathcal{R}_t = \sigma(M_t, X_{t+1}, X_{t+2}, \dots), \quad t = 1, 2, \dots$$

Proof. The exchangeability of (X_t) implies

$$\mathbb{E}[X_i | M_{t+1}] = \mathbb{E}[X_j | M_{t+1}], \quad i, j = 1, \dots, t+1,$$

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale with respect to the filtration:

$$\mathcal{R}_t = \sigma(M_t, X_{t+1}, X_{t+2}, \dots), \quad t = 1, 2, \dots$$

Proof. The exchangeability of (X_t) implies

$$\mathbb{E}[X_i | M_{t+1}] = \mathbb{E}[X_j | M_{t+1}], \quad i, j = 1, \dots, t+1,$$

hence, by independence (in fact, exchangeability suffices),

$$\mathbb{E}[X_i | \mathcal{R}_{t+1}] = \mathbb{E}[X_j | \mathcal{R}_{t+1}], \quad i, j = 1, \dots, t+1.$$

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale with respect to the filtration:

$$\mathcal{R}_t = \sigma(M_t, X_{t+1}, X_{t+2}, \dots), \quad t = 1, 2, \dots$$

Proof. The exchangeability of (X_t) implies

$$\mathbb{E}[X_i | M_{t+1}] = \mathbb{E}[X_j | M_{t+1}], \quad i, j = 1, \dots, t+1,$$

hence, by independence (in fact, exchangeability suffices),

$$\mathbb{E}[X_i | \mathcal{R}_{t+1}] = \mathbb{E}[X_j | \mathcal{R}_{t+1}], \quad i, j = 1, \dots, t+1.$$

Thus,

$$\mathbb{E}[M_t | \mathcal{R}_{t+1}] = \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}[X_i | \mathcal{R}_{t+1}]$$

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale with respect to the filtration:

$$\mathcal{R}_t = \sigma(M_t, X_{t+1}, X_{t+2}, \dots), \quad t = 1, 2, \dots$$

Proof. The exchangeability of (X_t) implies

$$\mathbb{E}[X_i|M_{t+1}] = \mathbb{E}[X_j|M_{t+1}], \quad i, j = 1, \dots, t+1,$$

hence, by independence (in fact, exchangeability suffices),

$$\mathbb{E}[X_i|\mathcal{R}_{t+1}] = \mathbb{E}[X_j|\mathcal{R}_{t+1}], \quad i, j = 1, \dots, t+1.$$

Thus,

$$\mathbb{E}[M_t|\mathcal{R}_{t+1}] = \frac{1}{t}\sum_{i=1}^{t}\mathbb{E}[X_i|\mathcal{R}_{t+1}] = \frac{1}{t+1}\sum_{i=1}^{t+1}\mathbb{E}[X_i|\mathcal{R}_{t+1}]$$

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale with respect to the filtration:

$$\mathcal{R}_t = \sigma(M_t, X_{t+1}, X_{t+2}, \dots), \quad t = 1, 2, \dots$$

Proof. The exchangeability of (X_t) implies

$$\mathbb{E}[X_i|M_{t+1}] = \mathbb{E}[X_j|M_{t+1}], \quad i, j = 1, \dots, t+1,$$

hence, by independence (in fact, exchangeability suffices),

$$\mathbb{E}[X_i|\mathcal{R}_{t+1}] = \mathbb{E}[X_j|\mathcal{R}_{t+1}], \quad i, j = 1, \dots, t+1.$$

Thus,

$$\mathbb{E}[M_t|\mathcal{R}_{t+1}] = \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}[X_i|\mathcal{R}_{t+1}] = \frac{1}{t+1}\sum_{i=1}^{t+1} \mathbb{E}[X_i|\mathcal{R}_{t+1}] = \mathbb{E}[M_{t+1}|\mathcal{R}_{t+1}]$$

Sample averages are reverse martingales

Claim. If $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then $M_t = \frac{1}{t}\sum_{i=1}^{t} X_i$ is a reverse martingale with respect to the filtration:

$$\mathcal{R}_t = \sigma(M_t, X_{t+1}, X_{t+2}, \dots), \quad t = 1, 2, \dots$$

Proof. The exchangeability of (X_t) implies

$$\mathbb{E}[X_i | M_{t+1}] = \mathbb{E}[X_j | M_{t+1}], \quad i, j = 1, \dots, t+1,$$

hence, by independence (in fact, exchangeability suffices),

$$\mathbb{E}[X_i | \mathcal{R}_{t+1}] = \mathbb{E}[X_j | \mathcal{R}_{t+1}], \quad i, j = 1, \dots, t+1.$$

Thus,

$$\mathbb{E}[M_t | \mathcal{R}_{t+1}] = \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}[X_i | \mathcal{R}_{t+1}] = \frac{1}{t+1}\sum_{i=1}^{t+1} \mathbb{E}[X_i | \mathcal{R}_{t+1}] = \mathbb{E}[M_{t+1} | \mathcal{R}_{t+1}] = M_{t+1}.$$

General (Sub, Super)Martingales

	Canonical Representation	Maximal Inequalities
Frwrd Supermart. Frwrd Mart. Frwrd Submart.	$X_0 + \sum_{i=1}^t X_i$, $X_i \in \mathcal{F}_i$, $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \leq 0$ $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] = 0$ $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \geq 0$	
Rev Supermart. Rev Mart. Rev Submart.	? $\frac{1}{t} \sum_{i=1}^t X_i$, $\mathbb{E}[X_i \mid \mathcal{R}_{t+1}] = M_{t+1}$?	

Nonnegative (Sub, Super)Martingales

Frwrd Supermart. Frwrd Mart. Frwrd Submart.	$X_0 \prod_{i=1}^t (1 + X_i)$, $X_i \in \mathcal{F}_i, X_i \geq -1$, $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \leq 0$ $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] = 0$ $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \geq 0$	
Rev Supermart. Rev Mart. Rev Submart.	?	

Nonnegative reverse martingales

We don't have a canonical form for nonnegative reverse martingales, but we can mimick one. e.g.: If $M_t = (1/t)\sum_{i=1}^t X_i$,

$$\exp(M_t) = \prod_{i=1}^t \exp(X_i/t) \quad \text{is a reverse submartingale.}$$

Nonnegative reverse martingales

We don't have a canonical form for nonnegative reverse martingales, but we can mimick one. e.g.: If $M_t = (1/t)\sum_{i=1}^t X_i$,

$$\exp(M_t) = \prod_{i=1}^{t} \exp(X_i/t) \quad \text{is a reverse submartingale.}$$

Indeed, if $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, and (M_t) is a reverse martingale w.r.t. $(\mathcal{R}_t)$, then $(\varphi(M_t))_{t=1}^{\infty}$ is a **reverse submartingale** w.r.t. $(\mathcal{R}_t)$.

Nonnegative reverse martingales

We don't have a canonical form for nonnegative reverse martingales, but we can mimick one. e.g.: If $M_t = (1/t)\sum_{i=1}^t X_i$,

$$\exp(M_t) = \prod_{i=1}^{t} \exp(X_i/t) \quad \text{is a reverse submartingale.}$$

Indeed, if $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, and (M_t) is a reverse martingale w.r.t. $(\mathcal{R}_t)$, then $(\varphi(M_t))_{t=1}^{\infty}$ is a **reverse submartingale** w.r.t. $(\mathcal{R}_t)$.

Proof. Follows from Jensen's inequality:

$$\mathbb{E}[\varphi(M_t)|\mathcal{R}_{t+1}] \geq \varphi(\mathbb{E}[M_t|\mathcal{R}_{t+1}]) = \varphi(M_{t+1}).$$

General (Sub, Super)Martingales

	Canonical Representation	Maximal Inequalities
Frwrd Supermart. Frwrd Mart. Frwrd Submart.	$X_0 + \sum_{i=1}^t X_i$, $X_i \in \mathcal{F}_i$; $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \leq 0$ $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] = 0$ $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \geq 0$	
Rev Supermart. Rev Mart. Rev Submart.	? $\frac{1}{t}\sum_{i=1}^t X_i$, $\mathbb{E}[X_i \mid \mathcal{R}_{t+1}] = M_{t+1}$?	

Nonnegative (Sub, Super)Martingales

Frwrd Supermart. Frwrd Mart. Frwrd Submart.	$X_0 \prod_{i=1}^t (1 + X_i)$, $X_i \in \mathcal{F}_i, X_i \geq -1$; $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \leq 0$ $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] = 0$ $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \geq 0$	
Rev Supermart. Rev Mart. Rev Submart.	? Example: $\prod_i \exp(X_i/t)$	

Maximal Inequalities for Reverse Martingales

Recall the prominent maximal inequalities for forward martingales

- **Ville's Inequality.** If (L_t) is a nonnegative forward supermartingale,

$$\mathbb{P}(\exists t \geq 1 : L_t \geq u) \leq \frac{\mathbb{E}[L_0]}{u}, \quad u > 0.$$

Recall the prominent maximal inequalities for forward martingales

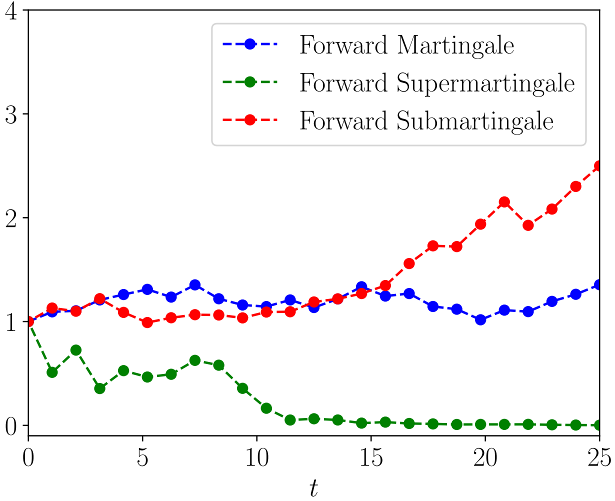
- **Ville's Inequality.** If (L_t) is a nonnegative forward supermartingale,

$$\mathbb{P}(\exists t \geq 1 : L_t \geq u) \leq \frac{\mathbb{E}[L_0]}{u}, \quad u > 0.$$

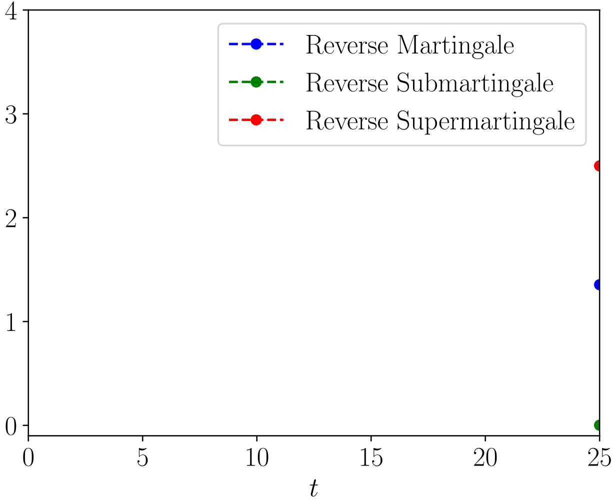
- **Doob's Submartingale Inequality.** If (L_t) is a nonnegative forward submartingale, then for all $T \geq 1$,

$$\mathbb{P}(\exists t \leq T : L_t \geq u) \leq \frac{\mathbb{E}[L_T]}{u}, \quad u > 0.$$

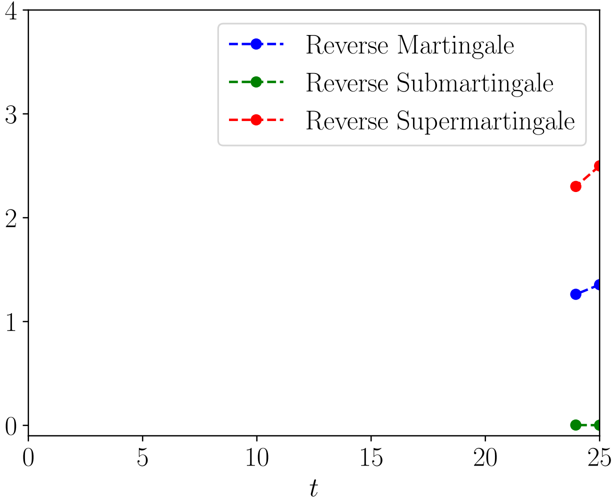
Ville's Inequality for Reverse Submartingales



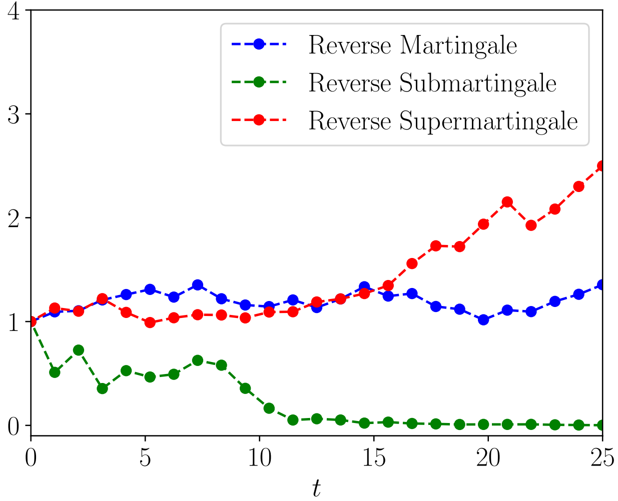
Ville's Inequality for Reverse Submartingales



Ville's Inequality for Reverse Submartingales



Ville's Inequality for Reverse Submartingales



Ville's Inequality for Reverse Submartingales

Let (M_t) be a nonnegative **reverse** submartingale with respect to $(\mathcal{R}_t)$.

Ville's Inequality for Reverse Submartingales

Let (M_t) be a nonnegative **reverse** submartingale with respect to $(\mathcal{R}_t)$.
For $T \geq 1$, set

$$L_t := M_{T-t+1}, \quad \mathcal{R}_t = \mathcal{F}_{T-t+1}, \quad 1 \leq t \leq T.$$

Ville′s Inequality for Reverse Submartingales

Let (M_t) be a nonnegative **reverse** submartingale with respect to $(\mathcal{R}_t)$.
For $T \geq 1$, set

$$L_t := M_{T-t+1}, \quad \mathcal{R}_t = \mathcal{F}_{T-t+1}, \quad 1 \leq t \leq T.$$

Then, $(L_t)_{t=1}^T$ is a nonnegative **forward** submartingale with respect to $(\mathcal{R}_t)_{t=1}^T$.

Ville's Inequality for Reverse Submartingales

Let (M_t) be a nonnegative **reverse** submartingale with respect to $(\mathcal{R}_t)$.
For $T \geq 1$, set

$$L_t := M_{T-t+1}, \quad \mathcal{R}_t = \mathcal{F}_{T-t+1}, \quad 1 \leq t \leq T.$$

Then, $(L_t)_{t=1}^T$ is a nonnegative **forward** submartingale with respect to $(\mathcal{R}_t)_{t=1}^T$. By Doob's submartingale inequality,

$$\mathbb{P}(\exists t \leq T : L_t \geq u) \leq \frac{\mathbb{E}[L_T]}{u}, \quad u > 0.$$

Ville′s Inequality for Reverse Submartingales

Let (M_t) be a nonnegative **reverse** submartingale with respect to $(\mathcal{R}_t)$.
For $T \geq 1$, set

$$L_t := M_{T-t+1}, \quad \mathcal{R}_t = \mathcal{F}_{T-t+1}, \quad 1 \leq t \leq T.$$

Then, $(L_t)_{t=1}^T$ is a nonnegative **forward** submartingale with respect to $(\mathcal{R}_t)_{t=1}^T$. By Doob's submartingale inequality,

$$\mathbb{P}(\exists t \leq T : L_t \geq u) \leq \frac{\mathbb{E}[L_T]}{u}, \quad u > 0.$$

This translates into

$$\mathbb{P}(\exists t \leq T : M_t \geq u) \leq \frac{\mathbb{E}[M_1]}{u}.$$

Ville's Inequality for Reverse Submartingales

Let (M_t) be a nonnegative **reverse** submartingale with respect to $(\mathcal{R}_t)$.
For $T \geq 1$, set

$$L_t := M_{T-t+1}, \quad \mathcal{R}_t = \mathcal{F}_{T-t+1}, \quad 1 \leq t \leq T.$$

Then, $(L_t)_{t=1}^T$ is a nonnegative **forward** submartingale with respect to $(\mathcal{R}_t)_{t=1}^T$. By Doob's submartingale inequality,

$$\mathbb{P}(\exists t \leq T : L_t \geq u) \leq \frac{\mathbb{E}[L_T]}{u}, \quad u > 0.$$

This translates into

$$\mathbb{P}(\exists t \leq T : M_t \geq u) \leq \frac{\mathbb{E}[M_1]}{u}.$$

Taking $T \to \infty$ leads to **Ville's inequality for reverse submartingales**:

$$\boxed{\mathbb{P}(\exists t \geq 1 : M_t \geq u) \leq \frac{\mathbb{E}[M_1]}{u}, \quad u > 0.}$$

General (Sub, Super)Martingales

	Canonical Representation	Maximal Inequalities
Frwrd Supermart.	$X_0 + \sum_{i=1}^t X_i$, $X_i \in \mathcal{F}_i$, $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \leq 0$	
Frwrd Mart.	$\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] = 0$	
Frwrd Submart.	$\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \geq 0$	
Rev Supermart.	?	
Rev Mart.	$\frac{1}{t}\sum_{i=1}^t X_i$, $\mathbb{E}[X_i \mid \mathcal{R}_{t+1}] = M_{t+1}$	
Rev Submart.	?	

Nonnegative (Sub, Super)Martingales

Frwrd Supermart.	$X_0 \prod_{i=1}^t (1 + X_i)$, $X_i \in \mathcal{F}_i, X_i \geq -1$, $\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \leq 0$	$\mathbb{P}(\exists t \geq 0 : L_t \geq u) \leq \mathbb{E}[L_0]/u$
Frwrd Mart.	$\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] = 0$	
Frwrd Submart.	$\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] \geq 0$	$\mathbb{P}(\exists t \leq T : L_t \geq u) \leq \mathbb{E}[L_T]/u$
Rev Supermart.		$\mathbb{P}(\exists t \leq T : M_t \geq u) \leq \mathbb{E}[M_T]/u$
Rev Mart.	?	
Rev Submart.	Example: $\prod_i \exp(X_i/t)$	$\mathbb{P}(\exists t \geq 1 : M_t \geq u) \leq \mathbb{E}[M_1]/u$

Measure-Valued Martingales and Exchangeable Filtrations

We know that if $(X_t)_{t=1}^{\infty}$ is a sequence of exchangeable RVs, then

$$\frac{1}{t}\sum_{i=1}^{t} X_i \quad \text{is a rev. martingale w.r.t. } \sigma\left(\frac{1}{t}\sum_{i=1}^{t} X_i, X_{t+1}, X_{t+2}, \dots\right).$$

We know that if $(X_t)_{t=1}^{\infty}$ is a sequence of exchangeable RVs, then

$$\frac{1}{t}\sum_{i=1}^{t} X_i \quad \text{is a rev. martingale w.r.t. } \sigma\left(\frac{1}{t}\sum_{i=1}^{t} X_i, X_{t+1}, X_{t+2}, \dots\right).$$

More generally, for any measurable function f,

$$\frac{1}{t}\sum_{i=1}^{t} f(X_i) \quad \text{is a rev. martingale w.r.t. } \sigma\left(\frac{1}{t}\sum_{i=1}^{t} f(X_i), f(X_{t+1}), f(X_{t+2}), \dots\right).$$

We know that if $(X_t)_{t=1}^{\infty}$ is a sequence of exchangeable RVs, then

$$\frac{1}{t}\sum_{i=1}^{t} X_i \quad \text{is a rev. martingale w.r.t. } \sigma\left(\frac{1}{t}\sum_{i=1}^{t} X_i, X_{t+1}, X_{t+2}, \dots\right).$$

More generally, for any measurable function f,

$$\frac{1}{t}\sum_{i=1}^{t} f(X_i) \quad \text{is a rev. martingale w.r.t. } \sigma\left(\frac{1}{t}\sum_{i=1}^{t} f(X_i), f(X_{t+1}), f(X_{t+2}), \dots\right).$$

Question: Does there exist a filtration $(\mathcal{E}_t)$ such that **for all** f, $\frac{1}{t}\sum_{i=1}^{t} f(X_i)$ is a reverse martingale with respect to $(\mathcal{E}_t)$?

We know that if $(X_t)_{t=1}^{\infty}$ is a sequence of exchangeable RVs, then

$$\frac{1}{t}\sum_{i=1}^{t} X_i \quad \text{is a rev. martingale w.r.t. } \sigma\left(\frac{1}{t}\sum_{i=1}^{t} X_i, X_{t+1}, X_{t+2}, \dots\right).$$

More generally, for any measurable function f,

$$\frac{1}{t}\sum_{i=1}^{t} f(X_i) \quad \text{is a rev. martingale w.r.t. } \sigma\left(\frac{1}{t}\sum_{i=1}^{t} f(X_i), f(X_{t+1}), f(X_{t+2}), \dots\right).$$

Question: Does there exist a filtration $(\mathcal{E}_t)$ such that **for all** f, $\frac{1}{t}\sum_{i=1}^{t} f(X_i)$ is a reverse martingale with respect to $(\mathcal{E}_t)$?
Answer: All we need is to ensure $\mathbb{E}[X_i|\mathcal{E}_t] = \mathbb{E}[X_j|\mathcal{E}_t]$ for all $i, j = 1, \dots, t$.

The Exchangeable Filtration

Given a sequence of random variables $(X_t)_{t=1}^{\infty}$, the exchangeable filtration $(\mathcal{E}_t)_{t=1}^{\infty}$ is defined by

$$\mathcal{E}_t = \sigma\left(\left\{h(X_1, X_2, \ldots, X_t) : \begin{array}{c} h \text{ is measurable and} \\ \text{permutation-symmetric} \end{array}\right\} \cup \{X_{t+1}, X_{t+2}, \ldots\}\right).$$

The Exchangeable Filtration

Given a sequence of random variables $(X_t)_{t=1}^{\infty}$, the exchangeable filtration $(\mathcal{E}_t)_{t=1}^{\infty}$ is defined by

$$\mathcal{E}_t = \sigma\left(\left\{h(X_1, X_2, \ldots, X_t) : \begin{array}{c} h \text{ is measurable and} \\ \text{permutation-symmetric} \end{array}\right\} \cup \{X_{t+1}, X_{t+2}, \ldots\}\right).$$

Equivalently, $\mathcal{E}_t$ is the set of events B whose indicator functions I_B are functions of $(X_t)_{t=1}^{\infty}$,

$$I_B = g(X_1, X_2, \ldots)$$

such that

$$g(X_1, X_2, \ldots) = g(X_{\tau(1)}, X_{\tau(2)}, \ldots, X_{\tau(t)}, X_{t+1}, X_{t+2}, \ldots),$$

for all permutations τ of $\{1, \ldots, t\}$.

The Exchangeable Filtration

Recall that $P_t = \frac{1}{t}\sum_{i=1}^{t}\delta_{X_i}$ denotes the empirical measure, and we have

$$\int f dP_t = \frac{1}{t}\sum_{i=1}^{t} f(X_i).$$

Theorem: Let $(X_t)_{t=1}^{\infty}$ be a sequence of exchangeable random variables. Then, for any bounded and measurable function f, $(\int f dP_t)$ is a reverse martingale with respect to $(\mathcal{E}_t)$. The converse holds true if (X_t) is stationary (Bladt 2019, Kallenberg 2005).

The Exchangeable Filtration

Recall that $P_t = \frac{1}{t}\sum_{i=1}^{t} \delta_{X_i}$ denotes the empirical measure, and we have

$$\int f dP_t = \frac{1}{t}\sum_{i=1}^{t} f(X_i).$$

Theorem: Let $(X_t)_{t=1}^{\infty}$ be a sequence of exchangeable random variables. Then, for any bounded and measurable function f, $(\int f dP_t)$ is a reverse martingale with respect to $(\mathcal{E}_t)$. The converse holds true if (X_t) is stationary (Bladt 2019, Kallenberg 2005).

We say that $(P_t)_{t=1}^{\infty}$ is a **measure-valued reverse martingale**.

A different perspective

If we believe $(\mathcal{E}_t)$ is the smallest filtration with respect to which (P_t) is a measure-valued reverse martingale, then we should heuristically have:

$$\mathcal{E}_t = \sigma(P_t, P_{t+1}, \dots), \quad t = 1, 2, \dots$$

A different perspective

If we believe $(\mathcal{E}_t)$ is the smallest filtration with respect to which (P_t) is a measure-valued reverse martingale, then we should heuristically have:

$$\mathcal{E}_t = \sigma(P_t, P_{t+1}, \dots), \quad t = 1, 2, \dots$$

- If we know P_t, then we know $X_1, \dots, X_t$ up to their ordering.

A different perspective

If we believe $(\mathcal{E}_t)$ is the smallest filtration with respect to which (P_t) is a measure-valued reverse martingale, then we should heuristically have:

$$\mathcal{E}_t = \sigma(P_t, P_{t+1}, \ldots), \quad t = 1, 2, \ldots$$

- If we know P_t, then we know $X_1, \ldots, X_t$ up to their ordering.
- If we know both P_{t+1} and P_t, then we know X_{t+1}.

A different perspective

If we believe $(\mathcal{E}_t)$ is the smallest filtration with respect to which (P_t) is a measure-valued reverse martingale, then we should heuristically have:

$$\mathcal{E}_t = \sigma(P_t, P_{t+1}, \dots), \quad t = 1, 2, \dots$$

- If we know P_t, then we know $X_1, \dots, X_t$ up to their ordering.
- If we know both P_{t+1} and P_t, then we know X_{t+1}.
- Altogether, $\mathcal{E}_t$ tells us the whole sequence $(X_t)_{t=1}^{\infty}$, except for the ordering of $X_1, \dots, X_t$.

A different perspective

If we believe $(\mathcal{E}_t)$ is the smallest filtration with respect to which (P_t) is a measure-valued reverse martingale, then we should heuristically have:

$$\mathcal{E}_t = \sigma(P_t, P_{t+1}, \dots), \quad t = 1, 2, \dots$$

- If we know P_t, then we know $X_1, \dots, X_t$ up to their ordering.
- If we know both P_{t+1} and P_t, then we know X_{t+1}.
- Altogether, $\mathcal{E}_t$ tells us the whole sequence $(X_t)_{t=1}^{\infty}$, except for the ordering of $X_1, \dots, X_t$.
- This is exactly the content of the exchangeable filtration!

The Hewitt-Savage 0-1 Law

Theorem: Assume $(X_t)_{t=1}^{\infty}$ is a sequence of i.i.d. random variables. Then, the exchangeable σ-algebra

$$\mathcal{E}_\infty = \bigcap_{t=1}^{\infty} \mathcal{E}_t$$

only contains events of probability zero or one.

Confidence Sequences for the One-Sample Problem

Recall the one-sample problem

Recall the one-sample problem

Given:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P,

Recall the one-sample problem

Given:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P,
- Empirical measure: $P_t = (1/t) \sum_{i=1}^{t} \delta_{X_i}$,

Recall the one-sample problem

Given:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P,
- Empirical measure: $P_t = (1/t) \sum_{i=1}^{t} \delta_{X_i}$,
- A general functional Φ. For instance, $\Phi(P) = D(P \| Q)$ for a fixed distribution Q and a divergence D.

Recall the one-sample problem

Given:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P,
- Empirical measure: $P_t = (1/t)\sum_{i=1}^{t} \delta_{X_i}$,
- A general functional Φ. For instance, $\Phi(P) = D(P \| Q)$ for a fixed distribution Q and a divergence D.

Goal: Derive (ℓ_t) and (u_t) such that

$$\mathbb{P}(\exists t \geq 1 : \Phi(P_t) - \ell_t \leq \Phi(P) \leq \Phi(P_t) + u_t) \geq 1 - \alpha.$$

Recall the one-sample problem

Given:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P,
- Empirical measure: $P_t = (1/t)\sum_{i=1}^{t} \delta_{X_i}$,
- A general functional Φ. For instance, $\Phi(P) = D(P\|Q)$ for a fixed distribution Q and a divergence D.

Goal: Derive (ℓ_t) and (u_t) such that

$$\mathbb{P}(\exists t \geq 1 : \Phi(P_t) - \ell_t \leq \Phi(P) \leq \Phi(P_t) + u_t) \geq 1 - \alpha.$$

(ℓ_t) and (u_t) will be obtained through separate approaches. We begin by deriving (ℓ_t).

Main idea

- (P_t) is a reverse measure-valued martingale.

Main idea

- (P_t) is a reverse measure-valued martingale.
- **Question**: In analogy to real-valued martingales, what is the "minimal" condition one can place on Φ to ensure that $\Phi(P_t)$ also has a martingale-type property?

Main idea

- (P_t) is a reverse measure-valued martingale.
- **Question**: In analogy to real-valued martingales, what is the "minimal" condition one can place on Φ to ensure that $\Phi(P_t)$ also has a martingale-type property?
- **Answer**: Convexity.

Main idea

- (P_t) is a reverse measure-valued martingale.
- **Question**: In analogy to real-valued martingales, what is the "minimal" condition one can place on Φ to ensure that $\Phi(P_t)$ also has a martingale-type property?
- **Answer**: Convexity.

Definition. The functional Φ is said to be convex if for all $\lambda \in [0, 1]$, and all probability distributions μ, ν,

$$\Phi(\lambda\mu + (1 - \lambda)\nu) \leq \lambda\Phi(\mu) + (1 - \lambda)\Phi(\nu).$$

Main idea

- (P_t) is a reverse measure-valued martingale.
- **Question**: In analogy to real-valued martingales, what is the "minimal" condition one can place on Φ to ensure that $\Phi(P_t)$ also has a martingale-type property?
- **Answer**: Convexity.

Definition. The functional Φ is said to be convex if for all $\lambda \in [0, 1]$, and all probability distributions μ, ν,

$$\Phi(\lambda\mu + (1 - \lambda)\nu) \leq \lambda\Phi(\mu) + (1 - \lambda)\Phi(\nu).$$

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Heuristic Proof. Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$.

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Heuristic Proof. Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Apply Jensen's inequality to (P_t),

$$\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi\big(\mathbb{E}[P_t|\mathcal{E}_{t+1}]\big)$$

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Heuristic Proof. Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Apply Jensen's inequality to (P_t),

$$\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi\big(\mathbb{E}[P_t|\mathcal{E}_{t+1}]\big) = \Phi(P_{t+1}),$$

because (P_t) is a reverse martingale.

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proper Proof (Sketch). Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$.

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proper Proof (Sketch). Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Form the following "leave-one-out" decomposition:

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proper Proof (Sketch). Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Form the following "leave-one-out" decomposition:

$$X_1, \underbrace{X_2, X_3, \ldots, X_{t-1}, X_t, X_{t+1}}, \ldots$$

$$P_t^1 = \frac{1}{t} \sum_{i=2}^{t+1} \delta_{X_i}$$

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proper Proof (Sketch). Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Form the following "leave-one-out" decomposition:

$$\underbrace{X_1}, X_2, \underbrace{X_3, \ldots, X_{t-1}, X_t, X_{t+1}}, \ldots$$

$$P_t^2 = \frac{1}{t} \sum_{i=1, i \neq 2}^{t+1} \delta_{X_i}$$

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proper Proof (Sketch). Need to show: $\mathbb{E}[\Phi(P_t) | \mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Form the following "leave-one-out" decomposition:

$$\underbrace{X_1, X_2}, X_3, \underbrace{X_4, \ldots, X_{t-1}, X_t, X_{t+1}}, \ldots$$

$$P_t^3 = \frac{1}{t} \sum_{i=1, i \neq 3}^{t+1} \delta_{X_i}, \text{ etc.}$$

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proper Proof (Sketch). Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Form the following "leave-one-out" decomposition:

$$P_{t+1} = \frac{1}{t+1}\sum_{k=1}^{t+1} P_t^k, \quad \text{where } P_t^k = \frac{1}{t}\sum_{\substack{i=1 \\ i\neq k}}^{t+1} \delta_{X_i}.$$

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proper Proof (Sketch). Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Form the following "leave-one-out" decomposition:

$$P_{t+1} = \frac{1}{t+1} \sum_{k=1}^{t+1} P_t^k, \quad \text{where } P_t^k = \frac{1}{t} \sum_{\substack{i=1 \\ i \neq k}}^{t+1} \delta_{X_i}.$$

By convexity, $\Phi(P_{t+1}) \leq \frac{1}{t+1} \sum_{k=1}^{t+1} \Phi(P_t^k)$,

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proper Proof (Sketch). Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Form the following "leave-one-out" decomposition:

$$P_{t+1} = \frac{1}{t+1} \sum_{k=1}^{t+1} P_t^k, \quad \text{where } P_t^k = \frac{1}{t} \sum_{\substack{i=1 \\ i \neq k}}^{t+1} \delta_{X_i}.$$

By convexity, $\Phi(P_{t+1}) \leq \frac{1}{t+1} \sum_{k=1}^{t+1} \Phi(P_t^k)$, thus,

$$\Phi(P_{t+1}) = \mathbb{E}[\Phi(P_{t+1})|\mathcal{E}_{t+1}] \leq \frac{1}{t+1} \sum_{k=1}^{t+1} \mathbb{E}[\Phi(P_t^k)|\mathcal{E}_{t+1}].$$

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proof Sketch. Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Have:

$$\Phi(P_{t+1}) = \mathbb{E}[\Phi(P_{t+1})|\mathcal{E}_{t+1}] \leq \frac{1}{t+1}\sum_{k=1}^{t+1}\mathbb{E}[\Phi(P_t^k)|\mathcal{E}_{t+1}].$$

Convex functionals evaluated at the empirical measure are reverse submartingales

Theorem. If Φ is convex, and $\Phi(P_t)$ has a finite first moment for all $t \geq 1$, then $(\Phi(P_t))$ *is a reverse submartingale* with respect to the exchangeable filtration $(\mathcal{E}_t)$ generated by (X_t).

Proof Sketch. Need to show: $\mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}] \geq \Phi(P_{t+1})$. Have:

$$\Phi(P_{t+1}) = \mathbb{E}[\Phi(P_{t+1})|\mathcal{E}_{t+1}] \leq \frac{1}{t+1}\sum_{k=1}^{t+1}\mathbb{E}[\Phi(P_t^k)|\mathcal{E}_{t+1}].$$

Now, similarly as for means,

$$\mathbb{E}[\Phi(P_t^k)|\mathcal{E}_{t+1}] = \mathbb{E}[\Phi(P_t)|\mathcal{E}_{t+1}], \quad k = 1, \ldots, t+1,$$

so we are done. $\square$

Some immediate consequences

- For any convex functional, the sequence

$$\big(\mathbb{E}[\Phi(P_t)]\big)_{t=1}^{\infty}$$

  is decreasing.

Some immediate consequences

- For any convex functional, the sequence

$$\big(\mathbb{E}[\Phi(P_t)]\big)_{t=1}^{\infty}$$

  is decreasing.
- As soon as $\Phi(P_t)$ is a consistent estimator of $\Phi(P)$, we obtain (assuming enough moments):

$$\mathbb{E}[\Phi(P_t)] \geq \mathbb{E}[\Phi(P)].$$

Some immediate consequences

- For any convex functional, the sequence

$$\big(\mathbb{E}[\Phi(P_t)]\big)_{t=1}^{\infty}$$

  is decreasing.
- As soon as $\Phi(P_t)$ is a consistent estimator of $\Phi(P)$, we obtain (assuming enough moments):

$$\mathbb{E}[\Phi(P_t)] \geq \mathbb{E}[\Phi(P)].$$

  Thus, plug-in estimators of convex functionals are upward biased.

Some immediate consequences

- For any convex functional, the sequence
$$\big(\mathbb{E}[\Phi(P_t)]\big)_{t=1}^{\infty}$$
is decreasing.
- As soon as $\Phi(P_t)$ is a consistent estimator of $\Phi(P)$, we obtain (assuming enough moments):
$$\mathbb{E}[\Phi(P_t)] \geq \mathbb{E}[\Phi(P)].$$
Thus, plug-in estimators of convex functionals are upward biased.
- Suppose $\Phi(P_t) \xrightarrow{p} \Phi(P)$ as $t \to \infty$. Then, $\Phi(P_t) \xrightarrow{a.s.} \Phi(P)$.
Proof.
  - By the reverse submartingale convergence theorem, $\Phi(P_t) \to Y$ for some $\mathcal{E}_\infty$-measurable Y.

Some immediate consequences

- For any convex functional, the sequence

$$\left(\mathbb{E}[\Phi(P_t)]\right)_{t=1}^{\infty}$$

  is decreasing.
- As soon as $\Phi(P_t)$ is a consistent estimator of $\Phi(P)$, we obtain (assuming enough moments):

$$\mathbb{E}[\Phi(P_t)] \geq \mathbb{E}[\Phi(P)].$$

  Thus, plug-in estimators of convex functionals are upward biased.
- Suppose $\Phi(P_t) \xrightarrow{p} \Phi(P)$ as $t \to \infty$. Then, $\Phi(P_t) \xrightarrow{a.s.} \Phi(P)$.
  Proof.
  - By the reverse submartingale convergence theorem, $\Phi(P_t) \to Y$ for some $\mathcal{E}_\infty$-measurable Y.
  - By Hewitt-Savage, Y is constant (a.s.).

Some immediate consequences

- For any convex functional, the sequence

$$\left(\mathbb{E}[\Phi(P_t)]\right)_{t=1}^{\infty}$$

  is decreasing.
- As soon as $\Phi(P_t)$ is a consistent estimator of $\Phi(P)$, we obtain (assuming enough moments):

$$\mathbb{E}[\Phi(P_t)] \geq \mathbb{E}[\Phi(P)].$$

  Thus, plug-in estimators of convex functionals are upward biased.
- Suppose $\Phi(P_t) \xrightarrow{p} \Phi(P)$ as $t \to \infty$. Then, $\Phi(P_t) \xrightarrow{a.s.} \Phi(P)$.
  Proof.
  - By the reverse submartingale convergence theorem, $\Phi(P_t) \to Y$ for some $\mathcal{E}_\infty$-measurable Y.
  - By Hewitt-Savage, Y is constant (a.s.).
  - Convergence in probability must imply that $Y = \Phi(P)$ a.s. $\square$

Lower confidence sequences for $\Phi(P)$

- $\Phi(P_t)$ is **not** a **nonnegative** reverse submartingale, so we can't use Ville's.

Lower confidence sequences for $\Phi(P)$

- $\Phi(P_t)$ is **not** a **nonnegative** reverse submartingale, so we can't use Ville's.
- If we think of $\Phi(P_t)$ as a mean, it is natural to instead work with $\exp(\Phi(P_t))$.

Lower confidence sequences for $\Phi(P)$

- $\Phi(P_t)$ is **not** a **nonnegative** reverse submartingale, so we can't use Ville's.
- If we think of $\Phi(P_t)$ as a mean, it is natural to instead work with $\exp(\Phi(P_t))$.

Assume $N_t := \Phi(P_t) - \Phi(P)$ has a finite MGF with a known upper bound:

$$\mathbb{E}\left[\exp(\lambda N_t)\right] \leq \exp(\psi_t(\lambda)), \quad \lambda \in [0, \lambda_{\max}).$$

Lower confidence sequences for $\Phi(P)$

- $\Phi(P_t)$ is **not** a **nonnegative** reverse submartingale, so we can't use Ville's.
- If we think of $\Phi(P_t)$ as a mean, it is natural to instead work with $\exp(\Phi(P_t))$.

Assume $N_t := \Phi(P_t) - \Phi(P)$ has a finite MGF with a known upper bound:

$$\mathbb{E}\left[\exp(\lambda N_t)\right] \leq \exp(\psi_t(\lambda)), \quad \lambda \in [0, \lambda_{\max}).$$

Then, for any λ,

$$L_t(\lambda) = \exp(\lambda N_t), \quad t = 1, 2, \ldots$$

is also reverse submartingale with respect to $(\mathcal{E}_t)$.

Lower confidence sequences via the Chernoff method

$$\begin{aligned}
\mathbb{P}(\exists t \geq 1 : N_t \geq u) &= \mathbb{P}\left(\exists t \geq 1 : \exp(\lambda N_t) \geq \exp(\lambda u)\right) \\
&\leq \frac{\mathbb{E}[\exp(\lambda N_1)]}{\exp(\lambda u)}, \quad \text{By Ville's inequality} \\
&\leq \exp(-\lambda u + \psi_1(\lambda)).
\end{aligned}$$

Lower confidence sequences via the Chernoff method

$$
\begin{aligned}
\mathbb{P}(\exists t \geq 1 : N_t \geq u) &= \mathbb{P}\left(\exists t \geq 1 : \exp(\lambda N_t) \geq \exp(\lambda u)\right) \\
&\leq \frac{\mathbb{E}[\exp(\lambda N_1)]}{\exp(\lambda u)}, \quad \text{By Ville's inequality} \\
&\leq \exp(-\lambda u + \psi_1(\lambda)).
\end{aligned}
$$

This held for any $\lambda \in [0, \lambda_{\max})$, thus,

$$
\mathbb{P}(\exists t \geq 1 : N_t \geq u) \leq \exp\left(- \sup_{\lambda \in [0, \lambda_{\max})} \{\lambda u - \psi_1(\lambda)\}\right) = \exp(-\psi_1^*(u)),
$$

where ψ_t^* is the convex conjugate of ψ_t. Inverting, we get:

$$
\boxed{\mathbb{P}\left\{\exists t \geq 1 : N_t \geq (\psi_1^*)^{-1}\big(\log(1/\alpha)\big)\right\} \leq \alpha.}
$$

But this confidence sequence has non-vanishing length!

Lower confidence sequences via the Chernoff method

$$\begin{aligned}\mathbb{P}(\exists t \geq 1 : N_t \geq u) &= \mathbb{P}\left(\exists t \geq 1 : \exp(\lambda N_t) \geq \exp(\lambda u)\right) \\ &\leq \frac{\mathbb{E}[\exp(\lambda N_1)]}{\exp(\lambda u)}, \quad \text{By Ville's inequality} \\ &\leq \exp(-\lambda u + \psi_1(\lambda)).\end{aligned}$$

This held for any $\lambda \in [0, \lambda_{\max})$, thus,

$$\mathbb{P}(\exists t \geq 1 : N_t \geq u) \leq \exp\left(-\sup_{\lambda \in [0,\lambda_{\max})} \left\{\lambda u - \psi_1(\lambda)\right\}\right) = \exp(-\psi_1^*(u)),$$

where ψ_t^* is the convex conjugate of ψ_t.

Lower confidence sequences via the Chernoff method

$$\begin{aligned}\mathbb{P}(\exists t \geq 1 : N_t \geq u) &= \mathbb{P}\left(\exists t \geq 1 : \exp(\lambda N_t) \geq \exp(\lambda u)\right) \\ &\leq \frac{\mathbb{E}[\exp(\lambda N_1)]}{\exp(\lambda u)}, \quad \text{By Ville's inequality} \\ &\leq \exp(-\lambda u + \psi_1(\lambda)).\end{aligned}$$

This held for any $\lambda \in [0, \lambda_{\max})$, thus,

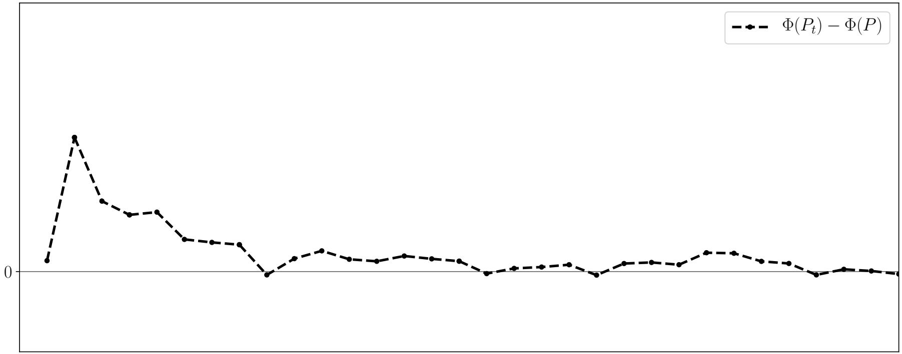
$$\mathbb{P}(\exists t \geq 1 : N_t \geq u) \leq \exp\left(-\sup_{\lambda \in [0,\lambda_{\max})} \left\{\lambda u - \psi_1(\lambda)\right\}\right) = \exp(-\psi_1^*(u)),$$

where ψ_t^* is the convex conjugate of ψ_t. Inverting, we get:

$$\boxed{\mathbb{P}\left\{\exists t \geq 1 : N_t \geq (\psi_1^*)^{-1}\big(\log(1/\alpha)\big)\right\} \leq \alpha.}$$

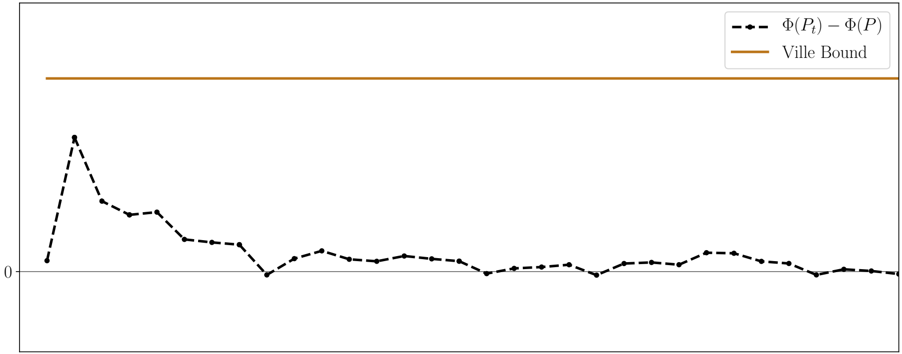
But this confidence sequence has non-vanishing length!

Stitched time-uniform Chernoff bounds



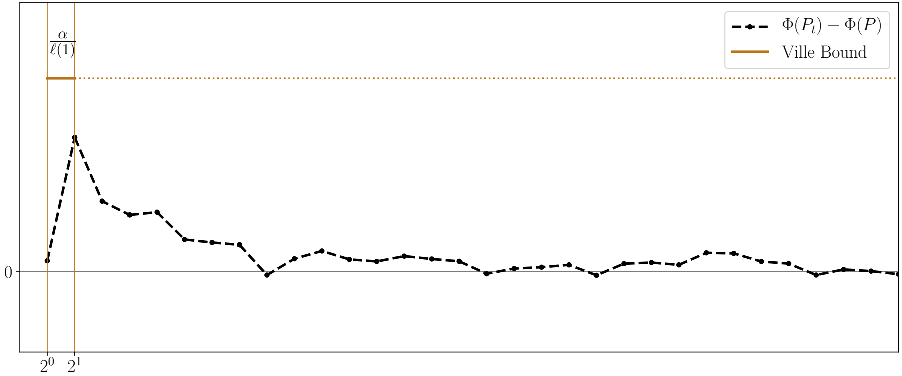
$$\mathbb{P}\Big\{\exists t \geq 1 : N_t \geq (\psi_1^*)^{-1}\big(\log(1/\alpha)\big)\Big\} \leq \alpha.$$

Stitched time-uniform Chernoff bounds



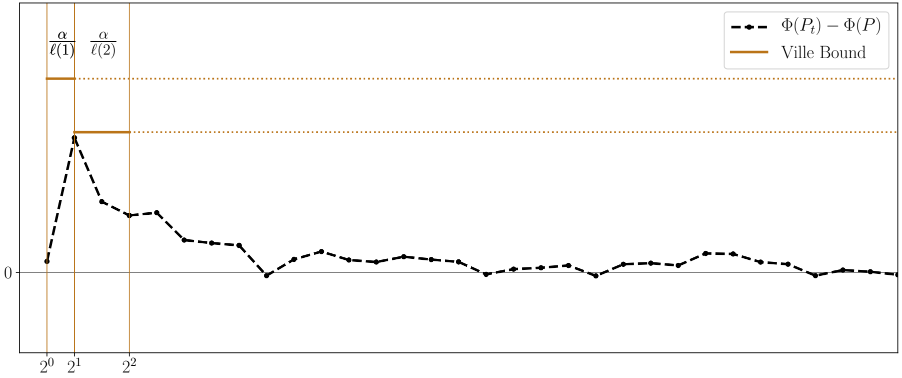
$$\mathbb{P}\Big\{\exists t \geq 1 : N_t \geq (\psi_1^*)^{-1}\big(\log(1/\alpha)\big)\Big\} \leq \alpha.$$

Stitched time-uniform Chernoff bounds



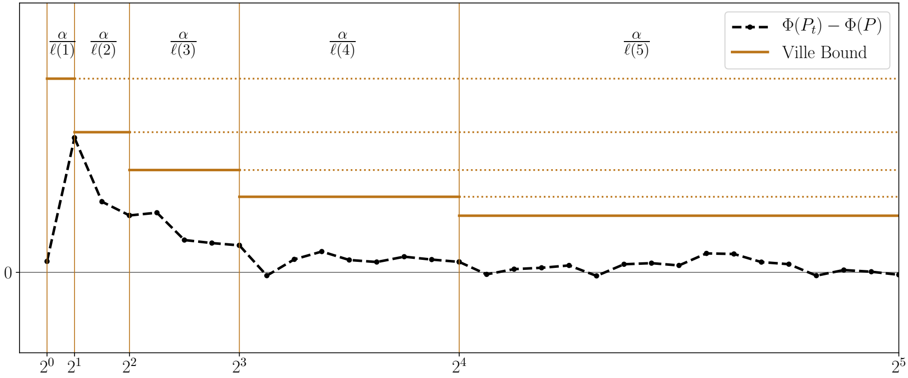
$$\mathbb{P}\Big\{\exists t \geq 1 : N_t \geq (\psi_1^*)^{-1}\big(\log(\ell(1)/\alpha)\big)\Big\} \leq \frac{\alpha}{\ell(1)}, \quad \ell(k) \geq 1.$$

Stitched time-uniform Chernoff bounds



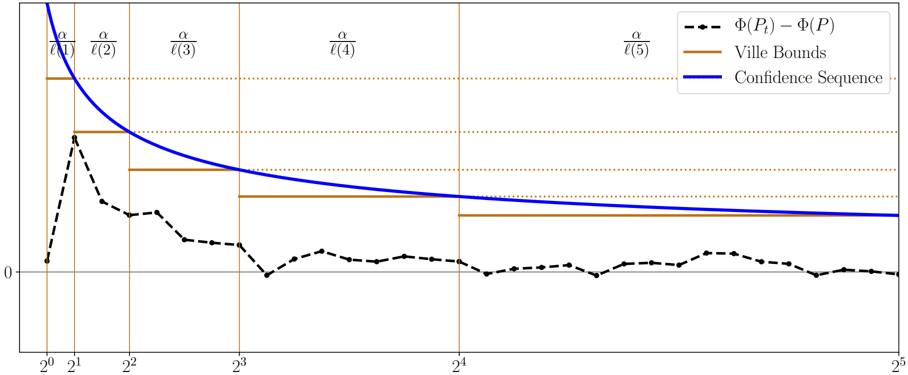
$$\mathbb{P}\left\{\exists t \geq 2^1 : N_t \geq (\psi_{2^1}^*)^{-1}\left(\log(\ell(2)/\alpha)\right)\right\} \leq \frac{\alpha}{\ell(2)}, \quad \ell(2) \geq 1.$$

Stitched time-uniform Chernoff bounds



$$\mathbb{P}\left\{\exists t \geq 2^k : N_t \geq ({\psi_{2^k}}^*)^{-1}\big(\log(\ell(k+1)/\alpha)\big)\right\} \leq \frac{\alpha}{\ell(k)}, \quad \ell(k) \geq 1, \sum_{k=1}^{\infty} \frac{1}{\ell(k)} = 1.$$

Stitched time-uniform Chernoff bounds



$$\mathbb{P}\left\{\exists t \geq 1 : N_t \geq (\psi^*_{t/2})^{-1}\big(\log(\ell(\log_2 t)/\alpha)\big)\right\}$$
$$\leq \sum_{k=0}^{\infty} \mathbb{P}\left\{\exists t \geq 2^k : N_t \geq ({\psi_{2^k}}^*)^{-1}\big(\log(\ell(k+1)/\alpha)\big)\right\} \leq \sum_{k=0}^{\infty} \frac{\alpha}{\ell(k+1)} = \alpha.$$

Stitched time-uniform Chernoff bounds

Theorem. Under the aforementioned moment assumption,

$$\mathbb{P}\Big\{\forall t \geq 1 : \Phi(P) \geq \Phi(P_t) - (\psi^*_{t/2})^{-1}\big(\log \ell(\log_2 t) + \log(1/\alpha)\big)\Big\} \geq 1 - \alpha.$$

Stitched time-uniform Chernoff bounds

Theorem. Under the aforementioned moment assumption,

$$\mathbb{P}\Big\{\forall t \geq 1 : \Phi(P) \geq \Phi(P_t) - (\psi^*_{t/2})^{-1}\big(\log \ell(\log_2 t) + \log(1/\alpha)\big)\Big\} \geq 1 - \alpha.$$

Example: $\sqrt{t}\Phi(P_t)$ is 1-sub-Gaussian, and if $\ell(k) \asymp k^2$,

$$\Phi(P) - \Phi(P_t) \geq -\,\mathbb{E}[\Phi(P_t) - \Phi(P)] - c\sqrt{\frac{\log\log t + \log(1/\alpha)}{t}}, \quad \text{with high probability,}$$

where $c > 0$ is a small constant depending on ℓ.

Stitched time-uniform Chernoff bounds

Theorem. Under the aforementioned moment assumption,

$$\mathbb{P}\Big\{\forall t \geq 1 : \Phi(P) \geq \Phi(P_t) - (\psi^*_{t/2})^{-1}\big(\log \ell(\log_2 t) + \log(1/\alpha)\big)\Big\} \geq 1 - \alpha.$$

Example: $\sqrt{t}\Phi(P_t)$ is 1-sub-Gaussian, and if $\ell(k) \asymp k^2$,

$$\Phi(P) - \Phi(P_t) \geq -\mathbb{E}[\Phi(P_t) - \Phi(P)] - c\sqrt{\frac{\log\log t + \log(1/\alpha)}{t}}, \quad \text{with high probability,}$$

where $c > 0$ is a small constant depending on ℓ.

Fixed-Time Chernoff Analogue:

$$\mathbb{P}\Big\{\forall t \geq 1 : \Phi(P) \geq \Phi(P_t) - (\psi^*_t)^{-1}\big(\log(1/\alpha)\big)\Big\} \geq 1 - \alpha.$$

Example 1: The Kolmogorov-Smirnov Statistic

Let $\Phi(Q) = \|F - G\|_\infty = \sup_{x \in \mathbb{R}} |F(x) - G(x)|$ where

- F is a fixed CDF
- G is the CDF of Q
- Let $F_t(x) = \frac{1}{t} \sum_{i=1}^{t} I(X_i \leq x)$ be the empirical CDF.

Example 1: The Kolmogorov-Smirnov Statistic

Let $\Phi(Q) = \|F - G\|_\infty = \sup_{x \in \mathbb{R}} |F(x) - G(x)|$ where

- F is a fixed CDF
- G is the CDF of Q
- Let $F_t(x) = \frac{1}{t} \sum_{i=1}^{t} I(X_i \leq x)$ be the empirical CDF.

The celebrated Dvoretzky-Kiefer-Wolfowitz (DKW) inequality states:

$$\mathbb{P}\left(\|F_t - F\|_\infty \leq \sqrt{\frac{\log(2/\alpha)}{2t}} \right) \geq 1 - \alpha.$$

Example 1: The Kolmogorov-Smirnov Statistic

Let $\Phi(Q) = \|F - G\|_\infty = \sup_{x \in \mathbb{R}} |F(x) - G(x)|$ where

- F is a fixed CDF
- G is the CDF of Q
- Let $F_t(x) = \frac{1}{t}\sum_{i=1}^{t} I(X_i \leq x)$ be the empirical CDF.

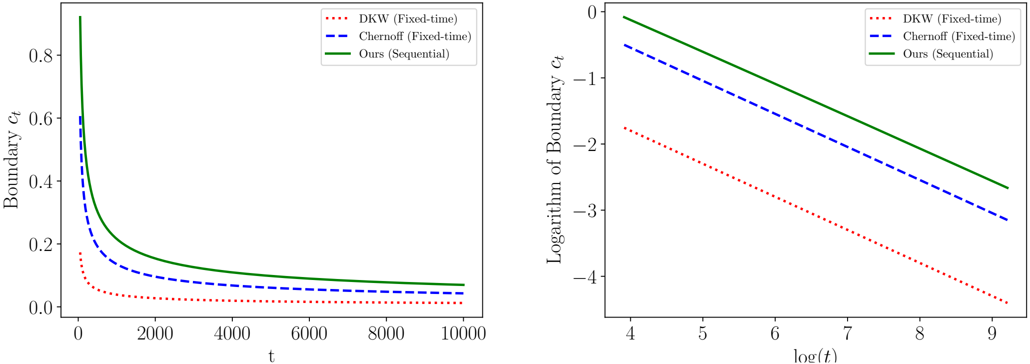
The celebrated Dvoretzky-Kiefer-Wolfowitz (DKW) inequality states:

$$\mathbb{P}\left(\|F_t - F\|_\infty \leq \sqrt{\frac{\log(2/\alpha)}{2t}} \right) \geq 1 - \alpha.$$

A slight modification of the previous bound leads to:

$$\mathbb{P}\left(\forall t \geq 3 : \|F_t - F\|_\infty \leq \sqrt{\frac{1.9}{t}} + 2.2\sqrt{\frac{1.2\log\log t + \log(4.5/\delta)}{t}} \right) \geq 1 - \alpha.$$

Example 1: The Kolmogorov-Smirnov Statistic



Where c_t is the critical value: $\mathbb{P}(\|F_t - F\|_\infty \geq c_t) \leq \alpha$ or $\mathbb{P}(\exists t \geq 1 : \|F_t - F\|_\infty \geq c_t) \leq \alpha$.

Example 2: The discrete Kullback-Leibler Divergence

Assume $P = \sum_{j=1}^{k} p_j \delta_{a_j}$ is supported on a finite set $\mathcal{X} = \{a_1, \ldots, a_k\}$ of size k.

Example 2: The discrete Kullback-Leibler Divergence

Assume $P = \sum_{j=1}^{k} p_j \delta_{a_j}$ is supported on a finite set $\mathcal{X} = \{a_1, \ldots, a_k\}$ of size k. Let,

$$C_j = P_t(\{a_j\}) = \sum_{i=1}^{t} I(X_i = a_j), \quad \text{so that } (C_1, \ldots, C_k) \sim \text{Multinomial}(t; p_1, \ldots, p_k).$$

Example 2: The discrete Kullback-Leibler Divergence

Assume $P = \sum_{j=1}^{k} p_j \delta_{a_j}$ is supported on a finite set $\mathcal{X} = \{a_1, \ldots, a_k\}$ of size k. Let,

$$C_j = P_t(\{a_j\}) = \sum_{i=1}^{t} I(X_i = a_j), \quad \text{so that } (C_1, \ldots, C_k) \sim \text{Multinomial}(t; p_1, \ldots, p_k).$$

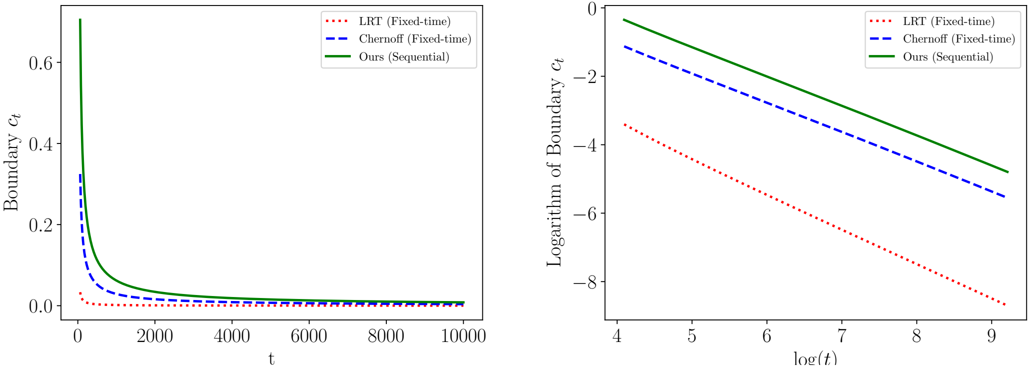
The MLE of $(p_1, \ldots, p_k)$ is $(\hat{p}_1, \ldots, \hat{p}_t) = (C_1/t, \ldots, C_k/t)$, and

$$\text{KL}(P_t \| P) = \sum_{j=1}^{k} \hat{p}_j \log\left(\frac{\hat{p}_j}{p_j}\right) = \frac{1}{t} \log\left(\frac{L_t(\hat{p})}{L_t(p)}\right)$$

is precisely the generalized log-likelihood ratio statistic up to rescaling! Thus:

$$2t\text{KL}(P_t \| P) \xrightarrow{d} \chi^2_{k-1}.$$

Example 2: The discrete Kullback-Leibler Divergence



Consider the special case where $\Phi(P) = \mathbb{E}_P[X] = \mu_0$, and P is supported on $[0, 1]$.

Consider the special case where $\Phi(P) = \mathbb{E}_P[X] = \mu_0$, and P is supported on $[0, 1]$.

Confidence Sequences via Stitching with Reverse Martingales	Confidence Sequences via Betting (Method of mixtures, etc.)

Consider the special case where $\Phi(P) = \mathbb{E}_P[X] = \mu_0$, and P is supported on $[0, 1]$.

Confidence Sequences via Stitching with Reverse Martingales	Confidence Sequences via Betting (Method of mixtures, etc.)
1. Construct a single process $\Phi(P_t)$, which is always a reverse submartingale.	

Consider the special case where $\Phi(P) = \mathbb{E}_P[X] = \mu_0$, and P is supported on $[0, 1]$.

Confidence Sequences via Stitching with Reverse Martingales	Confidence Sequences via Betting (Method of mixtures, etc.)
1. Construct a single process $\Phi(P_t)$, which is always a reverse submartingale. 2. Use Ville's inequality and the Chernoff method to obtain sets $\widetilde{C}_k$ such that $$\mathbb{P}(\forall t \geq 2^k : \mu_0 \in \widetilde{C}_k) \geq 1 - \frac{\alpha}{\ell(k+1)}.$$	

Consider the special case where $\Phi(P) = \mathbb{E}_P[X] = \mu_0$, and P is supported on $[0, 1]$.

Confidence Sequences via Stitching with Reverse Martingales	Confidence Sequences via Betting (Method of mixtures, etc.)
1. Construct a single process $\Phi(P_t)$, which is always a reverse submartingale. 2. Use Ville's inequality and the Chernoff method to obtain sets $\widetilde{C}_k$ such that $$\mathbb{P}(\forall t \geq 2^k : \mu_0 \in \widetilde{C}_k) \geq 1 - \frac{\alpha}{\ell(k+1)}.$$ 3. A confidence sequence for μ_0 is then: $$C_t = \widetilde{C}_k, \quad \text{where } t \in [2^k, 2^{k+1}).$$	

Consider the special case where $\Phi(P) = \mathbb{E}_P[X] = \mu_0$, and P is supported on $[0, 1]$.

Confidence Sequences via Stitching with Reverse Martingales	Confidence Sequences via Betting (Method of mixtures, etc.)
1. Construct a single process $\Phi(P_t)$, which is always a reverse submartingale.	1. Construct infinitely-many nonneg. processes $(M_t(\mu))$, indexed by $\mu \in \mathbb{R}$, s.t. only $(M_t(\mu_0))$ is a supermartingale. e.g.: Take $M_t(\mu) = \prod_{i=1}^{t}(1+\lambda_i(X_i-\mu))$.
2. Use Ville's inequality and the Chernoff method to obtain sets $\widetilde{C}_k$ such that $$\mathbb{P}(\forall t \geq 2^k : \mu_0 \in \widetilde{C}_k) \geq 1 - \frac{\alpha}{\ell(k+1)}.$$	
3. A confidence sequence for μ_0 is then: $$C_t = \widetilde{C}_k, \quad \text{where } t \in [2^k, 2^{k+1}).$$	

Consider the special case where $\Phi(P) = \mathbb{E}_P[X] = \mu_0$, and P is supported on $[0, 1]$.

Confidence Sequences via Stitching with Reverse Martingales	Confidence Sequences via Betting (Method of mixtures, etc.)
1. Construct a single process $\Phi(P_t)$, which is always a reverse submartingale.	1. Construct infinitely-many nonneg. processes $(M_t(\mu))$, indexed by $\mu \in \mathbb{R}$, s.t. only $(M_t(\mu_0))$ is a supermartingale. e.g.: Take $M_t(\mu) = \prod_{i=1}^{t}(1+\lambda_i(X_i-\mu))$.
2. Use Ville's inequality and the Chernoff method to obtain sets $\widetilde{C}_k$ such that $$\mathbb{P}(\forall t \geq 2^k : \mu_0 \in \widetilde{C}_k) \geq 1 - \frac{\alpha}{\ell(k+1)}.$$	2. By Ville's inequality, $$\mathbb{P}(\forall t \geq 1 : M_t(\mu_0) \leq 1/\alpha) \geq 1 - \alpha.$$
3. A confidence sequence for μ_0 is then: $$C_t = \widetilde{C}_k, \quad \text{where } t \in [2^k, 2^{k+1}).$$	

Consider the special case where $\Phi(P) = \mathbb{E}_P[X] = \mu_0$, and P is supported on $[0, 1]$.

Confidence Sequences via Stitching with Reverse Martingales	Confidence Sequences via Betting (Method of mixtures, etc.)
1. Construct a single process $\Phi(P_t)$, which is always a reverse submartingale.	1. Construct infinitely-many nonneg. processes $(M_t(\mu))$, indexed by $\mu \in \mathbb{R}$, s.t. only $(M_t(\mu_0))$ is a supermartingale. e.g.: Take $M_t(\mu) = \prod_{i=1}^{t}(1+\lambda_i(X_i-\mu))$.
2. Use Ville's inequality and the Chernoff method to obtain sets $\widetilde{C}_k$ such that $$\mathbb{P}(\forall t \geq 2^k : \mu_0 \in \widetilde{C}_k) \geq 1 - \frac{\alpha}{\ell(k+1)}.$$	2. By Ville's inequality, $$\mathbb{P}(\forall t \geq 1 : M_t(\mu_0) \leq 1/\alpha) \geq 1 - \alpha.$$
3. A confidence sequence for μ_0 is then: $$C_t = \widetilde{C}_k, \quad \text{where } t \in [2^k, 2^{k+1}).$$	3. A confidence sequence for μ_0 is then: $$C_t = \{\mu \in [0,1] : M_t(\mu) \leq 1/\alpha\}.$$

Upper confidence sequences

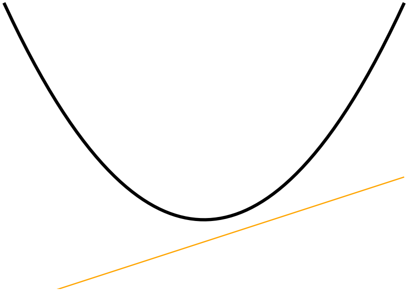
We have found (ℓ_t) such that:

$$\forall t \geq 1 : \quad -\ell(t) \leq \Phi(P) - \Phi(P_t), \quad \text{with high probability.}$$

Can we use the same strategy to find (u_t) such that

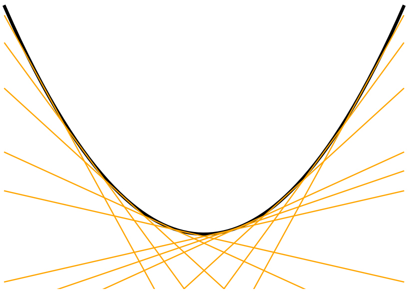
$$\forall t \geq 1 : \quad \Phi(P) - \Phi(P_t) \leq u_t, \quad \text{with high probability?}$$

Upper confidence sequences via affine minorants



Any convex function f admits an affine minorant.

Upper confidence sequences via affine minorants



Any convex function f admits an affine minorant. In fact, under some conditions,

$$f(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - f^*(\lambda) \right\} = f^{**}(x).$$

Upper confidence sequences via affine minorants

$$f(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - f^*(\lambda) \right\} = f^{**}(x).$$

With some care, the same holds true for convex functionals.

Upper confidence sequences via affine minorants

$$f(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - f^*(\lambda) \right\} = f^{**}(x).$$

With some care, the same holds true for convex functionals.

- We say a functional L is linear if $L(\lambda P + Q) = \lambda L(P) + L(Q)$ for all $\lambda \in \mathbb{R}$ and all distributions P, Q.

Upper confidence sequences via affine minorants

$$f(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - f^*(\lambda) \right\} = f^{**}(x).$$

With some care, the same holds true for convex functionals.

- We say a functional L is linear if $L(\lambda P + Q) = \lambda L(P) + L(Q)$ for all $\lambda \in \mathbb{R}$ and all distributions P, Q.

Upper confidence sequences via affine minorants

$$f(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - f^*(\lambda) \right\} = f^{**}(x).$$

With some care, the same holds true for convex functionals.

- We say a functional L is linear if $L(\lambda P + Q) = \lambda L(P) + L(Q)$ for all $\lambda \in \mathbb{R}$ and all distributions P, Q.

Heuristically, we thus expect:

$$\Phi(P) = \sup_{L} \left\{ L(P) - \Phi^*(L) \right\}.$$

Assume L_0 achieves the supremum. Then,

$$\Phi(P_t) - \Phi(P) \geq \left\{ L_0(P_t) - \Phi^*(L_0) \right\} - \left\{ L_0(P) - \Phi^*(L_0) \right\} = L_0(P_t) - L_0(P)$$

Upper confidence sequences via affine minorants

$$\Phi(P_t) - \Phi(P) \geq \Big\{ L_0(P_t) - \Phi^*(L_0) \Big\} - \Big\{ L_0(P) - \Phi^*(L_0) \Big\} = L_0(P_t) - L_0(P)$$

Upper confidence sequences via affine minorants

$$\Phi(P_t) - \Phi(P) \geq \left\{L_0(P_t) - \Phi^*(L_0)\right\} - \left\{L_0(P) - \Phi^*(L_0)\right\} = L_0(P_t) - L_0(P)$$

L_0 is linear, thus

$$L_0(P_t) = L_0\left(\frac{1}{t}\sum_{i=1}^{t}\delta_{X_i}\right) = \frac{1}{t}\sum_{i=1}^{t} L_0(\delta_{X_i}).$$

This is a sample average, hence a reverse martingale!

Upper confidence sequences via affine minorants

$$\Phi(P_t) - \Phi(P) \geq \Big\{L_0(P_t) - \Phi^*(L_0)\Big\} - \Big\{L_0(P) - \Phi^*(L_0)\Big\} = L_0(P_t) - L_0(P)$$

L_0 is linear, thus

$$L_0(P_t) = L_0\left(\frac{1}{t}\sum_{i=1}^{t}\delta_{X_i}\right) = \frac{1}{t}\sum_{i=1}^{t}L_0(\delta_{X_i}).$$

This is a sample average, hence a reverse martingale! Thus:

$$\Phi(P_t) - \Phi(P) \geq \frac{1}{t}\sum_{i=1}^{t}\big[L_0(\delta_{X_i}) - L_0(P)\big].$$

In many well-known cases, the summand has mean zero. A high probability bound follows.

Example: Maximum Mean Discrepancy

Given a reproducing kernel K bounded by B, recall:

$$D^2(P\|Q) = \mathbb{E}_{X,X'\sim P}[K(X,X')] + \mathbb{E}_{Y,Y'\sim Q}\big[K(Y,Y')\big] - 2\mathbb{E}_{\substack{X\sim P\\ Y\sim Q}}\big[K(X,Y)\big].$$

Example: Maximum Mean Discrepancy

Given a reproducing kernel K bounded by B, recall:

$$D^2(P\|Q) = \mathbb{E}_{X,X'\sim P}[K(X,X')] + \mathbb{E}_{Y,Y'\sim Q}\left[K(Y,Y')\right] - 2\mathbb{E}_{\substack{X\sim P\\ Y\sim Q}}\left[K(X,Y)\right].$$

We obtain,

$$\mathbb{P}(\forall t \geq 1 : -\ell_t \leq D(P\|Q) - D(P_t\|Q) \leq u_t) \geq 1-\alpha.$$

where,

$$\ell_t = 4\sqrt{\frac{B}{t}\left[1.2\log\log t + \log(9/\alpha)\right]} + 2\sqrt{\frac{2B}{t}}$$

$$u_t = 2\sqrt{\frac{B}{t}\left[1.2\log\log t + \log(9/\alpha)\right]}.$$

Example: Maximum Mean Discrepancy

Given a reproducing kernel K bounded by B, recall:

$$D^2(P\|Q) = \mathbb{E}_{X,X'\sim P}[K(X,X')] + \mathbb{E}_{Y,Y'\sim Q}\left[K(Y,Y')\right] - 2\mathbb{E}_{\substack{X\sim P\\ Y\sim Q}}\left[K(X,Y)\right].$$

We obtain,

$$\mathbb{P}(\forall t \geq 1 : -\ell_t \leq D(P\|Q) - D(P_t\|Q) \leq u_t) \geq 1-\alpha.$$

where,

$$\ell_t = 4\sqrt{\frac{B}{t}\left[1.2\log\log t + \log(9/\alpha)\right]} + 2\sqrt{\frac{2B}{t}}$$

$$u_t = 2\sqrt{\frac{B}{t}\left[1.2\log\log t + \log(9/\alpha)\right]}.$$

- Can the bias term be removed?

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

- Canonical example: $\Phi(P) = \mathrm{Var}(X)$ when $h(x, y) = (x - y)^2/2$.

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

- Canonical example: $\Phi(P) = \mathrm{Var}(X)$ when $h(x, y) = (x - y)^2/2$.

V-Statistic Estimator: $\Phi(P_t) = \frac{1}{t^2} \sum_{i,j=1}^{t} h(X_i, X_j)$ is called a V-Statistic.

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

- Canonical example: $\Phi(P) = \mathrm{Var}(X)$ when $h(x, y) = (x - y)^2/2$.

V-Statistic Estimator: $\Phi(P_t) = \frac{1}{t^2} \sum_{i,j=1}^{t} h(X_i, X_j)$ is called a V-Statistic.

Proposition. If h is a reproducing kernel, then $\sqrt{\Phi}$, and hence also Φ, are convex functionals. In particular, $(\Phi(P_t))$ and $(\sqrt{\Phi(P_t)})$ are reverse submartingales with respect to $(\mathcal{E}_t)$.

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

- Canonical example: $\Phi(P) = \mathrm{Var}(X)$ when $h(x, y) = (x - y)^2/2$.

V-Statistic Estimator: $\Phi(P_t) = \frac{1}{t^2} \sum_{i,j=1}^{t} h(X_i, X_j)$ is called a V-Statistic.

Proposition. If h is a reproducing kernel, then $\sqrt{\Phi}$, and hence also Φ, are convex functionals. In particular, $(\Phi(P_t))$ and $(\sqrt{\Phi(P_t)})$ are reverse submartingales with respect to $(\mathcal{E}_t)$.

- Confidence sequences can thus be derived for $\Phi(P)$ based on $\Phi(P_t)$.

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

- Canonical example: $\Phi(P) = \mathrm{Var}(X)$ when $h(x, y) = (x - y)^2/2$.

V-Statistic Estimator: $\Phi(P_t) = \frac{1}{t^2}\sum_{i,j=1}^{t} h(X_i, X_j)$ is called a V-Statistic.

Proposition. If h is a reproducing kernel, then $\sqrt{\Phi}$, and hence also Φ, are convex functionals. In particular, $(\Phi(P_t))$ and $(\sqrt{\Phi(P_t)})$ are reverse submartingales with respect to $(\mathcal{E}_t)$.

- Confidence sequences can thus be derived for $\Phi(P)$ based on $\Phi(P_t)$.
- Φ is convex, so $\Phi(P_t)$ is upwards biased. It turns out that our methods extend to the unbiased U-Statistic estimator or $\Phi(P)$.

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

Unbiased (U-Statistic) estimator of $\Phi(P)$:

$$U_t = \frac{2}{t(t-1)} \sum_{1 \le i < j \le t} h(X_i, X_j).$$

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

Unbiased (U-Statistic) estimator of $\Phi(P)$:

$$U_t = \frac{2}{t(t-1)} \sum_{1 \le i < j \le t} h(X_i, X_j).$$

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

Unbiased (U-Statistic) estimator of $\Phi(P)$:

$$U_t = \frac{2}{t(t-1)} \sum_{1 \le i < j \le t} h(X_i, X_j).$$

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

- This fact was established by Berk (1966) with respect to a distinct filtration.

Further Examples: V-Statistics and U-Statistics

Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a symmetric function. Set

$$\Phi(P) = \mathbb{E}[h(X, X')], \quad \text{where } X, X' \sim P \text{ are independent.}$$

Unbiased (U-Statistic) estimator of $\Phi(P)$:

$$U_t = \frac{2}{t(t-1)} \sum_{1 \le i < j \le t} h(X_i, X_j).$$

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

- This fact was established by Berk (1966) with respect to a distinct filtration.
- Our past approach for lower confidence sequences can now be adapted to obtain both lower and upper confidence sequences based on U_t.

Further Examples: V-Statistics and U-Statistics

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

Proof Sketch. Either proceed as before, or note that:

$$U_t = \mathbb{E}[U_t|\mathcal{E}_t] = \frac{2}{t(t-1)} \sum_{1 \le i < j \le t} \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t].$$

Further Examples: V-Statistics and U-Statistics

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

Proof Sketch. Either proceed as before, or note that:

$$U_t = \mathbb{E}[U_t|\mathcal{E}_t] = \frac{2}{t(t-1)} \sum_{1 \le i < j \le t} \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t].$$

By symmetry,

$$\mathbb{E}[h(X_1, X_2)|\mathcal{E}_t] = \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t], \quad \forall i \neq j.$$

Further Examples: V-Statistics and U-Statistics

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

Proof Sketch. Either proceed as before, or note that:

$$U_t = \mathbb{E}[U_t|\mathcal{E}_t] = \frac{2}{t(t-1)} \sum_{1\leq i<j\leq t} \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t].$$

By symmetry,

$$\mathbb{E}[h(X_1, X_2)|\mathcal{E}_t] = \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t], \quad \forall i \neq j.$$

Therefore,

$$U_t = \mathbb{E}[h(X_1, X_2)|\mathcal{E}_t].$$

Further Examples: V-Statistics and U-Statistics

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

Proof Sketch. Either proceed as before, or note that:

$$U_t = \mathbb{E}[U_t|\mathcal{E}_t] = \frac{2}{t(t-1)} \sum_{1 \le i < j \le t} \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t].$$

By symmetry,

$$\mathbb{E}[h(X_1, X_2)|\mathcal{E}_t] = \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t], \quad \forall i \neq j.$$

Therefore,

$$U_t = \mathbb{E}[h(X_1, X_2)|\mathcal{E}_t].$$

whence,

$$\mathbb{E}[U_t|\mathcal{E}_{t+1}] = \mathbb{E}\big\{\mathbb{E}[h(X_1, X_2)|\mathcal{E}_t]|\mathcal{E}_{t+1}\big\}$$

Further Examples: V-Statistics and U-Statistics

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

Proof Sketch. Either proceed as before, or note that:

$$U_t = \mathbb{E}[U_t|\mathcal{E}_t] = \frac{2}{t(t-1)} \sum_{1 \le i < j \le t} \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t].$$

By symmetry,

$$\mathbb{E}[h(X_1, X_2)|\mathcal{E}_t] = \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t], \quad \forall i \neq j.$$

Therefore,

$$U_t = \mathbb{E}[h(X_1, X_2)|\mathcal{E}_t].$$

whence,

$$\mathbb{E}[U_t|\mathcal{E}_{t+1}] = \mathbb{E}\big\{\mathbb{E}[h(X_1, X_2)|\mathcal{E}_t]|\mathcal{E}_{t+1}\big\} = \mathbb{E}[h(X_1, X_2)|\mathcal{E}_{t+1}]$$

Further Examples: V-Statistics and U-Statistics

Proposition. $(U_t)_{t=1}^{\infty}$ is a **reverse martingale** with respect to $(\mathcal{E}_t)_{t=1}^{\infty}$.

Proof Sketch. Either proceed as before, or note that:

$$U_t = \mathbb{E}[U_t|\mathcal{E}_t] = \frac{2}{t(t-1)} \sum_{1 \leq i < j \leq t} \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t].$$

By symmetry,

$$\mathbb{E}[h(X_1, X_2)|\mathcal{E}_t] = \mathbb{E}[h(X_i, X_j)|\mathcal{E}_t], \quad \forall i \neq j.$$

Therefore,

$$U_t = \mathbb{E}[h(X_1, X_2)|\mathcal{E}_t].$$

whence,

$$\mathbb{E}[U_t|\mathcal{E}_{t+1}] = \mathbb{E}\{\mathbb{E}[h(X_1, X_2)|\mathcal{E}_t]|\mathcal{E}_{t+1}\} = \mathbb{E}[h(X_1, X_2)|\mathcal{E}_{t+1}] = U_{t+1}.$$

□

Discussion on the Choice of Filtration

- $(\Phi(P_t))$ is a reverse submartingale with respect to $(\mathcal{E}_t)$.

Discussion on the Choice of Filtration

- $(\Phi(P_t))$ is a reverse submartingale with respect to $(\mathcal{E}_t)$.
- $(\Phi(P_t))$ is also lower-bounded by a reverse martingale with respect to $(\mathcal{E}_t)$.

Discussion on the Choice of Filtration

- $(\Phi(P_t))$ is a reverse submartingale with respect to $(\mathcal{E}_t)$.
- $(\Phi(P_t))$ is also lower-bounded by a reverse martingale with respect to $(\mathcal{E}_t)$.
- But $(\Phi(P_t))$ is also adapted to the **canonical forward filtration**,

$$\mathcal{D}_t = \sigma(X_1, \ldots, X_t), \quad t = 1, 2, \ldots$$

Discussion on the Choice of Filtration

- $(\Phi(P_t))$ is a reverse submartingale with respect to $(\mathcal{E}_t)$.
- $(\Phi(P_t))$ is also lower-bounded by a reverse martingale with respect to $(\mathcal{E}_t)$.
- But $(\Phi(P_t))$ is also adapted to the **canonical forward filtration**,

$$\mathcal{D}_t = \sigma(X_1, \ldots, X_t), \quad t = 1, 2, \ldots$$

- $(\mathcal{E}_t)$ was a useful **design tool**, but we will ultimately do statistical inference with respect $(\mathcal{D}_t)$.

Discussion on the Choice of Filtration

- $(\Phi(P_t))$ is a reverse submartingale with respect to $(\mathcal{E}_t)$.
- $(\Phi(P_t))$ is also lower-bounded by a reverse martingale with respect to $(\mathcal{E}_t)$.
- But $(\Phi(P_t))$ is also adapted to the **canonical forward filtration**,

$$\mathcal{D}_t = \sigma(X_1, \ldots, X_t), \quad t = 1, 2, \ldots$$

- $(\mathcal{E}_t)$ was a useful **design tool**, but we will ultimately do statistical inference with respect $(\mathcal{D}_t)$. e.g.:

$$\mathbb{P}(\Phi(P) \in C_\tau) \geq 1 - \alpha, \quad \text{for any stopping time } \tau \textbf{ with respect to } (\mathcal{D}_t).$$

Confidence Sequences for the Two-Sample Problem and Partially-Ordered Martingales

Recall the two-sample problem

Recall the two-sample problem

Given:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P, and $(Y_s)_{s=1}^{\infty}$ from an unknown distribution Q.

Recall the two-sample problem

Given:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P, and $(Y_s)_{s=1}^{\infty}$ from an unknown distribution Q.
- Empirical measures: $P_t = (1/t)\sum_{i=1}^{t} \delta_{X_i}$, $Q_s = (1/s)\sum_{j=1}^{s} \delta_{Y_j}$.

Recall the two-sample problem

<u>Given</u>:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P, and $(Y_s)_{s=1}^{\infty}$ from an unknown distribution Q.
- Empirical measures: $P_t = (1/t)\sum_{i=1}^{t}\delta_{X_i}$, $Q_s = (1/s)\sum_{j=1}^{s}\delta_{Y_j}$.
- A bivariate functional Ψ. For instance, $\Psi(P,Q) = D(P\|Q)$ for a divergence D.

Recall the two-sample problem

<u>Given</u>:

- Observations $(X_t)_{t=1}^{\infty}$ i.i.d. from an unknown distribution P, and $(Y_s)_{s=1}^{\infty}$ from an unknown distribution Q.
- Empirical measures: $P_t = (1/t) \sum_{i=1}^{t} \delta_{X_i}$, $Q_s = (1/s) \sum_{j=1}^{s} \delta_{Y_j}$.
- A bivariate functional Ψ. For instance, $\Psi(P, Q) = D(P \| Q)$ for a divergence D.

What is a reasonable definition of two-sample confidence sequence?

Suppose we wish to test whether (X_t) and (Y_s) have the same mean. Start by testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1. Assume $X_i, Y_i \in [0, 1]$.

Suppose we wish to test whether (X_t) and (Y_s) have the same mean. Start by testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1. Assume $X_i, Y_i \in [0,1]$.

Protocol: Testing two means μ_0, μ_1.

$\mathcal{K}_0 = 1$
for $n = 1, 2, \ldots$ **do**
  Skeptic announces
   $\lambda^{(i)} \in \left[-\frac{1}{1-\mu_i}, \frac{1}{\mu_i}\right], \; i = 0, 1.$
  Realities announce $I_n \in \{0, 1\}$.
  if $I_n = 0$ **then**
    Reality 0 announces X
    $\mathcal{K}_n = \mathcal{K}_{n-1}\big(1 + \lambda^{(0)}(X - \mu_0)\big)$
  else
    Reality 1 announces Y
    $\mathcal{K}_n = \mathcal{K}_{n-1}\big(1 + \lambda^{(1)}(Y - \mu_1)\big)$
  end
end

Suppose we wish to test whether (X_t) and (Y_s) have the same mean. Start by testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1. Assume $X_i, Y_i \in [0,1]$.

Protocol: Testing two means μ_0, μ_1.

$\mathcal{K}_0 = 1$
for $n = 1, 2, \ldots$ **do**
  Skeptic announces
  $\lambda^{(i)} \in \left[-\frac{1}{1-\mu_i}, \frac{1}{\mu_i}\right], \; i = 0, 1.$
  Realities announce $I_n \in \{0,1\}$.
  if $I_n = 0$ **then**
    $\lambda^{(0)}_{T(n)} = \lambda^{(0)}$
    Reality 0 announces $X_{T(n)}$
    $\mathcal{K}_n = \mathcal{K}_{n-1}\left(1 + \lambda^{(0)}_{T(n)}(X_{T(n)} - \mu_0)\right)$
  else
    $\lambda^{(1)}_{S(n)} = \lambda^{(1)}$
    Reality 1 announces $Y_{S(n)}$
    $\mathcal{K}_n = \mathcal{K}_{n-1}\left(1 + \lambda^{(1)}_{S(n)}(Y_{S(n)} - \mu_1)\right)$
  end
end

$$T(n) = \sum_{j=1}^{n}(1 - I_j), \quad S(n) = \sum_{j=1}^{n} I_j.$$

Bet at time n: $(1 - I_n)\lambda^{(0)}_{T(n)} + I_n \lambda^{(1)}_{S(n)}$.

Suppose we wish to test whether (X_t) and (Y_s) have the same mean. Start by testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1. Assume $X_i, Y_i \in [0,1]$.

Protocol: Testing two means μ_0, μ_1.

$\mathcal{K}_0 = 1$

for $n = 1, 2, \ldots$ **do**

Skeptic announces

$\lambda^{(i)} \in \left[-\frac{1}{1-\mu_i}, \frac{1}{\mu_i}\right], \ i = 0, 1.$

Realities announce $I_n \in \{0, 1\}$.

if $I_n = 0$ **then**

$\lambda^{(0)}_{T(n)} = \lambda^{(0)}$

Reality 0 announces $X_{T(n)}$

$\mathcal{K}_n = \mathcal{K}_{n-1}\big(1 + \lambda^{(0)}_{T(n)}(X_{T(n)} - \mu_0)\big)$

else

$\lambda^{(1)}_{S(n)} = \lambda^{(1)}$

Reality 1 announces $Y_{S(n)}$

$\mathcal{K}_n = \mathcal{K}_{n-1}\big(1 + \lambda^{(1)}_{S(n)}(Y_{S(n)} - \mu_1)\big)$

end

end

$$T(n) = \sum_{j=1}^{n}(1 - I_j), \quad S(n) = \sum_{j=1}^{n} I_j.$$

Bet at time n: $(1 - I_n)\lambda^{(0)}_{T(n)} + I_n\lambda^{(1)}_{S(n)}$.

Obs. at time n: $(1 - I_n)X_{T(n)} + I_n Y_{S(n)}$.

Suppose we wish to test whether (X_t) and (Y_s) have the same mean. Start by testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1. Assume $X_i, Y_i \in [0,1]$.

Protocol: Testing two means μ_0, μ_1.

$\mathcal{K}_0 = 1$

for $n = 1, 2, \ldots$ **do**

  Skeptic announces

  $\lambda^{(i)} \in \left[-\frac{1}{1-\mu_i}, \frac{1}{\mu_i}\right], \; i = 0, 1.$

  Realities announce $I_n \in \{0, 1\}$.

  if $I_n = 0$ **then**

    $\lambda^{(0)}_{T(n)} = \lambda^{(0)}$

    Reality 0 announces $X_{T(n)}$

    $\mathcal{K}_n = \mathcal{K}_{n-1}\big(1 + \lambda^{(0)}_{T(n)}(X_{T(n)} - \mu_0)\big)$

  else

    $\lambda^{(1)}_{S(n)} = \lambda^{(1)}$

    Reality 1 announces $Y_{S(n)}$

    $\mathcal{K}_n = \mathcal{K}_{n-1}\big(1 + \lambda^{(1)}_{S(n)}(Y_{S(n)} - \mu_1)\big)$

  end

end

$$T(n) = \sum_{j=1}^{n}(1 - I_j), \quad S(n) = \sum_{j=1}^{n} I_j.$$

Bet at time n: $(1 - I_n)\lambda^{(0)}_{T(n)} + I_n\lambda^{(1)}_{S(n)}$.

Obs. at time n: $(1 - I_n)X_{T(n)} + I_n Y_{S(n)}$.

Two-Sample Capital Process:

$$\mathcal{K}_n = \prod_{j=1}^{n}\Big[\big(1 + \lambda^{(0)}_{T(j)}(X_{T(j)} - \mu_0)\big)^{1-I_j} \cdot \big(1 + \lambda^{(1)}_{S(j)}(Y_{S(j)} - \mu_1)\big)^{I_j}\Big]$$

The two-sample capital process

$$\mathcal{K}_n = \prod_{j=1}^{n} \left(1 + \lambda_{T(j)}^{(0)}(X_{T(j)} - \mu_0)\right)^{1-I_j} \left(1 + \lambda_{S(j)}^{(1)}(Y_{S(j)} - \mu_1)\right)^{I_j}$$

- If (I_n) is a deterministic sequence, then $\mathcal{K}_n$ is a nonnegative martingale w.r.t.:

$$\mathcal{D}_t = \sigma(Z_1, Z_2, \dots), \quad \text{where } Z_n = (1 - I_n)X_{T(n)} + I_n Y_{S(n)}.$$

The two-sample capital process

$$\mathcal{K}_n = \prod_{j=1}^{n} \left(1 + \lambda^{(0)}_{T(j)}(X_{T(j)} - \mu_0)\right)^{1-I_j} \left(1 + \lambda^{(1)}_{S(j)}(Y_{S(j)} - \mu_1)\right)^{I_j}$$

- If (I_n) is a deterministic sequence, then $\mathcal{K}_n$ is a nonnegative martingale w.r.t.:

$$\mathcal{D}_t = \sigma(Z_1, Z_2, \dots), \quad \text{where } Z_n = (1 - I_n)X_{T(n)} + I_n Y_{S(n)}.$$

- The same holds true if (I_n) is predictable (with some care about the filtration).

The two-sample capital process

$$\mathcal{K}_n = \prod_{j=1}^{n} \left(1 + \lambda_{T(j)}^{(0)}(X_{T(j)} - \mu_0)\right)^{1-I_j} \left(1 + \lambda_{S(j)}^{(1)}(Y_{S(j)} - \mu_1)\right)^{I_j}$$

- If (I_n) is a deterministic sequence, then $\mathcal{K}_n$ is a nonnegative martingale w.r.t.:

$$\mathcal{D}_t = \sigma(Z_1, Z_2, \dots), \quad \text{where } Z_n = (1 - I_n)X_{T(n)} + I_n Y_{S(n)}.$$

- The same holds true if (I_n) is predictable (with some care about the filtration).
- We deduce $\mathbb{P}(\exists n \geq 1 : \mathcal{K}_n(\mu_0, \mu_1) \geq 1/\alpha) \leq \alpha$, thus

The two-sample capital process

$$\mathcal{K}_n = \prod_{j=1}^{n} \left(1 + \lambda^{(0)}_{T(j)}(X_{T(j)} - \mu_0)\right)^{1-I_j} \left(1 + \lambda^{(1)}_{S(j)}(Y_{S(j)} - \mu_1)\right)^{I_j}$$

- If (I_n) is a deterministic sequence, then $\mathcal{K}_n$ is a nonnegative martingale w.r.t.:

$$\mathcal{D}_t = \sigma(Z_1, Z_2, \dots), \quad \text{where } Z_n = (1 - I_n)X_{T(n)} + I_n Y_{S(n)}.$$

- The same holds true if (I_n) is predictable (with some care about the filtration).
- We deduce $\mathbb{P}(\exists n \geq 1 : \mathcal{K}_n(\mu_0, \mu_1) \geq 1/\alpha) \leq \alpha$, thus

$$C_n = \{(\mu_0, \mu_1) : \mathcal{K}_n(\mu_0, \mu_1) \leq 1/\alpha\}$$

  is a $1 - \alpha$ confidence sequence for $(\mathbb{E}_P[X], \mathbb{E}_Q[Y])$.

The two-sample capital process

$$\mathcal{K}_n = \prod_{j=1}^{n} \left(1 + \lambda^{(0)}_{T(j)}(X_{T(j)} - \mu_0)\right)^{1-I_j} \left(1 + \lambda^{(1)}_{S(j)}(Y_{S(j)} - \mu_1)\right)^{I_j}$$

- If (I_n) is a deterministic sequence, then $\mathcal{K}_n$ is a nonnegative martingale w.r.t.:

$$\mathcal{D}_t = \sigma(Z_1, Z_2, \dots), \quad \text{where } Z_n = (1 - I_n)X_{T(n)} + I_n Y_{S(n)}.$$

- The same holds true if (I_n) is predictable (with some care about the filtration).
- We deduce $\mathbb{P}(\exists n \geq 1 : \mathcal{K}_n(\mu_0, \mu_1) \geq 1/\alpha) \leq \alpha$, thus

$$C_n = \{(\mu_0, \mu_1) : \mathcal{K}_n(\mu_0, \mu_1) \leq 1/\alpha\}$$

  is a $1 - \alpha$ confidence sequence for $(\mathbb{E}_P[X], \mathbb{E}_Q[Y])$.

- $I\left(\exists \mu : (\mu, \mu) \in C_n\right)$ is a sequential $1 - \alpha$ test for $H_0 : \mu_0 = \mu_1$.

Generalization to convex functionals

Let Ψ be a bivariate convex functional. If (I_n) is deterministic,

$$(\Psi(P_{T(n)}, Q_{S(n)}))_{n=1}^{\infty} \quad \text{is a reverse submartingale w.r.t. exch. filtr. of } (Z_n)_{n=1}^{\infty}.$$

Similarly as before, we thus get $[\ell_n, u_n]$ s.t.:

$$\mathbb{P}\left(\forall n \geq 1 : -\ell_n \leq \Psi(P, Q) - \Psi(P_{T(n)} \| Q_{S(n)}) \leq u_n\right) \geq 1 - \alpha.$$

- We conjecture this is also true if (I_n) depends on $Z_1, \ldots, Z_{n-1}$, but not on the order of $X_1, \ldots, X_{T(n)}$ and $Y_1, \ldots, Y_{S(n)}$.

Generalization to convex functionals

Let Ψ be a bivariate convex functional. If (I_n) is deterministic,

$$(\Psi(P_{T(n)}, Q_{S(n)}))_{n=1}^{\infty} \quad \text{is a reverse submartingale w.r.t. exch. filtr. of } (Z_n)_{n=1}^{\infty}.$$

Similarly as before, we thus get $[\ell_n, u_n]$ s.t.:

$$\mathbb{P}\left(\forall n \geq 1 : -\ell_n \leq \Psi(P, Q) - \Psi(P_{T(n)} \| Q_{S(n)}) \leq u_n\right) \geq 1 - \alpha.$$

- We conjecture this is also true if (I_n) depends on $Z_1, \ldots, Z_{n-1}$, but not on the order of $X_1, \ldots, X_{T(n)}$ and $Y_1, \ldots, Y_{S(n)}$.
- What if we want (I_n) to depend arbitrarily on the data? Including at time n?

Again, consider testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1.

Again, consider testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1.

Protocol: Separately testing two means.

$\mathcal{K}_0 = 2$
$\mathcal{K}_0^{(0)} = \mathcal{K}_0^{(1)} = 1$
for $t = 1, 2, \ldots$ **do**
  Skeptic 0 announces $\lambda_t^{(0)} \in \left[\frac{-1}{1-\mu_0}, \frac{1}{\mu_0}\right]$
  Reality 0 announces X_t
  $\mathcal{K}_t^{(0)} = \mathcal{K}_{t-1}^{(0)}\big(1 + \lambda_t^{(0)}(X_t - \mu_0)\big)$
end
for $s = 1, 2, \ldots$ **do**
  Skeptic 1 announces $\lambda_s^{(1)} \in \left[\frac{-1}{1-\mu_1}, \frac{1}{\mu_1}\right]$
  Reality 1 announces X_t
  $\mathcal{K}_s^{(1)} = \mathcal{K}_{s-1}^{(1)}\big(1 + \lambda_s^{(1)}(Y_s - \mu_1)\big)$
end
Capital at times (t, s): $\mathcal{K}_{ts} = \mathcal{K}_t^{(0)} + \mathcal{K}_s^{(1)}$

Again, consider testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1.

Protocol: Separately testing two means.

$\mathcal{K}_0 = 2$

$\mathcal{K}_0^{(0)} = \mathcal{K}_0^{(1)} = 1$

for $t = 1, 2, \ldots$ **do**

  Skeptic 0 announces $\lambda_t^{(0)} \in \left[\frac{-1}{1-\mu_0}, \frac{1}{\mu_0}\right]$

  Reality 0 announces X_t

  $\mathcal{K}_t^{(0)} = \mathcal{K}_{t-1}^{(0)}\big(1 + \lambda_t^{(0)}(X_t - \mu_0)\big)$

end

for $s = 1, 2, \ldots$ **do**

  Skeptic 1 announces $\lambda_s^{(1)} \in \left[\frac{-1}{1-\mu_1}, \frac{1}{\mu_1}\right]$

  Reality 1 announces X_t

  $\mathcal{K}_s^{(1)} = \mathcal{K}_{s-1}^{(1)}\big(1 + \lambda_s^{(1)}(Y_s - \mu_1)\big)$

end

Capital at times (t, s): $\mathcal{K}_{ts} = \mathcal{K}_t^{(0)} + \mathcal{K}_s^{(1)}$

The Partially-Ordered Capital Process:

$$\mathcal{K}_{ts} = \prod_{i=1}^{t}(1 + \lambda_t^{(0)}(X_t - \mu_0)) + \prod_{j=1}^{s}\big(1 + \lambda_s^{(1)}(Y_s - \mu_1)\big).$$

Again, consider testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1.

Protocol: Separately testing two means.

$\mathcal{K}_0 = 2$
$\mathcal{K}_0^{(0)} = \mathcal{K}_0^{(1)} = 1$
for $t = 1, 2, \ldots$ **do**
  Skeptic 0 announces $\lambda_t^{(0)} \in \left[\frac{-1}{1-\mu_0}, \frac{1}{\mu_0}\right]$
  Reality 0 announces X_t
  $\mathcal{K}_t^{(0)} = \mathcal{K}_{t-1}^{(0)}\big(1 + \lambda_t^{(0)}(X_t - \mu_0)\big)$
end
for $s = 1, 2, \ldots$ **do**
  Skeptic 1 announces $\lambda_s^{(1)} \in \left[\frac{-1}{1-\mu_1}, \frac{1}{\mu_1}\right]$
  Reality 1 announces X_t
  $\mathcal{K}_s^{(1)} = \mathcal{K}_{s-1}^{(1)}\big(1 + \lambda_s^{(1)}(Y_s - \mu_1)\big)$
end
Capital at times (t, s): $\mathcal{K}_{ts} = \mathcal{K}_t^{(0)} + \mathcal{K}_s^{(1)}$

The Partially-Ordered Capital Process:

$$\mathcal{K}_{ts} = \prod_{i=1}^{t}(1 + \lambda_t^{(0)}(X_t - \mu_0)) + \prod_{j=1}^{s}\big(1 + \lambda_s^{(1)}(Y_s - \mu_1)\big).$$

- $(\mathcal{K}_t^0)$ and $(\mathcal{K}_s^1)$ are martingales, so

$$\mathbb{P}(\exists t \geq 1 : \mathcal{K}_t^{(0)} \geq \mathcal{K}_0^{(0)}/\alpha) \leq \alpha.$$

$$\mathbb{P}(\exists s \geq 1 : \mathcal{K}_s^{(1)} \geq \mathcal{K}_0^{(1)}/\alpha) \leq \alpha.$$

Again, consider testing if (X_t) has mean μ_0 and (Y_s) has mean μ_1.

Protocol: Separately testing two means.

$\mathcal{K}_0 = 2$

$\mathcal{K}_0^{(0)} = \mathcal{K}_0^{(1)} = 1$

for $t = 1, 2, \ldots$ **do**

- Skeptic 0 announces $\lambda_t^{(0)} \in \left[\frac{-1}{1-\mu_0}, \frac{1}{\mu_0}\right]$
- Reality 0 announces X_t
- $\mathcal{K}_t^{(0)} = \mathcal{K}_{t-1}^{(0)}\left(1 + \lambda_t^{(0)}(X_t - \mu_0)\right)$

end

for $s = 1, 2, \ldots$ **do**

- Skeptic 1 announces $\lambda_s^{(1)} \in \left[\frac{-1}{1-\mu_1}, \frac{1}{\mu_1}\right]$
- Reality 1 announces X_t
- $\mathcal{K}_s^{(1)} = \mathcal{K}_{s-1}^{(1)}\left(1 + \lambda_s^{(1)}(Y_s - \mu_1)\right)$

end

Capital at times (t, s): $\mathcal{K}_{ts} = \mathcal{K}_t^{(0)} + \mathcal{K}_s^{(1)}$

The Partially-Ordered Capital Process:

$$\mathcal{K}_{ts} = \prod_{i=1}^{t}(1 + \lambda_t^{(0)}(X_t - \mu_0)) + \prod_{j=1}^{s}\left(1 + \lambda_s^{(1)}(Y_s - \mu_1)\right).$$

- $(\mathcal{K}_t^0)$ and $(\mathcal{K}_s^1)$ are martingales, so

$$\mathbb{P}(\exists t \geq 1 : \mathcal{K}_t^{(0)} \geq \mathcal{K}_0^{(0)}/\alpha) \leq \alpha.$$

$$\mathbb{P}(\exists s \geq 1 : \mathcal{K}_s^{(1)} \geq \mathcal{K}_0^{(1)}/\alpha) \leq \alpha.$$

hence,

$$\mathbb{P}(\exists t, s \geq 1 : \mathcal{K}_{ts} \geq \mathcal{K}_{00}/\alpha) \leq 2\alpha.$$

Partially-ordered processes become more important for functionals other than means.

Partially-ordered processes become more important for functionals other than means. Two approaches:

- n**-Uniform Confidence Sequence.** Assume the ordering $T(n)$ and $S(n)$ are known, or data-independent. Then, can obtain ℓ_n, u_n such that:

$$\mathbb{P}(\forall n \geq 1 : -\ell_n \leq \Psi(P, Q) - \Psi(P_{T(n)}, Q_{S(n)}) \leq u_n) \geq 1 - \alpha.$$

- (t, s)**-Uniform Confidence Sequence.** Assume the data comes in any data-dependent order. Then, we would like to find ℓ_{ts} and u_{ts} such that

$$\mathbb{P}(\forall t, s \geq 1 : -\ell_{ts} \leq \Psi(P, Q) - \Psi(P_t, Q_s) \leq u_{ts}) \geq 1 - \alpha.$$

We will need partially-ordered martingales to handle the latter.

Partially-ordered forward filtrations and martingales

A partially-ordered forward filtration is a sequence $(\mathcal{F}_{ts})_{t,s=1}^{\infty}$ of σ-algebras which is increasing with respect to the partial order on $\mathbb{N}^2$, i.e.:

$$\mathcal{F}_{ts} \subseteq \mathcal{F}_{(t+1)s}, \quad \mathcal{F}_{ts} \subseteq \mathcal{F}_{t(s+1)}, \quad t, s = 1, 2, \ldots$$

Partially-ordered forward filtrations and martingales

A partially-ordered forward filtration is a sequence $(\mathcal{F}_{ts})_{t,s=1}^{\infty}$ of σ-algebras which is increasing with respect to the partial order on $\mathbb{N}^2$, i.e.:

$$\mathcal{F}_{ts} \subseteq \mathcal{F}_{(t+1)s}, \quad \mathcal{F}_{ts} \subseteq \mathcal{F}_{t(s+1)}, \quad t, s = 1, 2, \ldots$$

A partially-ordered forward martingale with respect to $(\mathcal{F}_{ts})_{t=1}^{\infty}$ is a process $(M_{ts})_{t,s=1}^{\infty}$ such that:

- $M_{ts} \in \mathcal{F}_{ts}$ for all $t, s \geq 1$.

Partially-ordered forward filtrations and martingales

A partially-ordered forward filtration is a sequence $(\mathcal{F}_{ts})_{t,s=1}^{\infty}$ of σ-algebras which is increasing with respect to the partial order on $\mathbb{N}^2$, i.e.:

$$\mathcal{F}_{ts} \subseteq \mathcal{F}_{(t+1)s}, \quad \mathcal{F}_{ts} \subseteq \mathcal{F}_{t(s+1)}, \quad t, s = 1, 2, \ldots$$

A partially-ordered forward martingale with respect to $(\mathcal{F}_{ts})_{t=1}^{\infty}$ is a process $(M_{ts})_{t,s=1}^{\infty}$ such that:

- $M_{ts} \in \mathcal{F}_{ts}$ for all $t, s \geq 1$.
- We have,

$$\mathbb{E}[M_{(t+1)s} | \mathcal{F}_{ts}] = M_{ts}, \quad \mathbb{E}[M_{t(s+1)} | \mathcal{F}_{ts}] = M_{ts}, \quad t, s = 0, 1, \ldots.$$

Partially-ordered forward filtrations and martingales

A partially-ordered forward filtration is a sequence $(\mathcal{F}_{ts})_{t,s=1}^{\infty}$ of σ-algebras which is increasing with respect to the partial order on $\mathbb{N}^2$, i.e.:

$$\mathcal{F}_{ts} \subseteq \mathcal{F}_{(t+1)s}, \quad \mathcal{F}_{ts} \subseteq \mathcal{F}_{t(s+1)}, \quad t, s = 1, 2, \ldots$$

A partially-ordered forward martingale with respect to $(\mathcal{F}_{ts})_{t=1}^{\infty}$ is a process $(M_{ts})_{t,s=1}^{\infty}$ such that:

- $M_{ts} \in \mathcal{F}_{ts}$ for all $t, s \geq 1$.
- We have,

$$\mathbb{E}[M_{(t+1)s} | \mathcal{F}_{ts}] = M_{ts}, \quad \mathbb{E}[M_{t(s+1)} | \mathcal{F}_{ts}] = M_{ts}, \quad t, s = 0, 1, \ldots.$$

Replace by $\leq$ for supermartingales and $\geq$ for submartingales.

Partially-ordered reverse filtrations and martingales

A partially-ordered reverse filtration is a sequence $(\mathcal{F}_{ts})_{t,s=1}^{\infty}$ of σ-algebras which is decreasing with respect to the partial order on $\mathbb{N}^2$, i.e.:

$$\mathcal{F}_{ts} \supseteq \mathcal{F}_{(t+1)s}, \quad \mathcal{F}_{ts} \supseteq \mathcal{F}_{t(s+1)}, \quad t, s = 1, 2, \ldots$$

A partially-ordered reverse martingale with respect to $(\mathcal{F}_{ts})_{t=1}^{\infty}$ is a process $(M_{ts})_{t,s=1}^{\infty}$ such that:

- $M_{ts} \in \mathcal{F}_{ts}$ for all $t, s \geq 1$.
- We have,

$$\mathbb{E}[M_{ts} | \mathcal{F}_{(t+1)s}] = M_{(t+1)s}, \quad \mathbb{E}[M_{ts} | \mathcal{F}_{t(s+1)}] = M_{t(s+1)}, \quad t, s = 0, 1, \ldots.$$

Replace by $\leq$ for supermartingales and $\geq$ for submartingales.

Ville′s Inequality does not hold for partially-ordered martingales

Thoerem (Cairoli, 1970). There exists a forward martingale (M_{ts}) and $u > 0$ such that

$$\mathbb{P}(\exists t, s \geq 0 : M_{ts} \geq u) > \frac{\mathbb{E}[M_{00}]}{u}.$$

Ville′s Inequality does not hold for partially-ordered martingales

Thoerem (Cairoli, 1970). There exists a forward martingale (M_{ts}) and $u > 0$ such that

$$\mathbb{P}(\exists t, s \geq 0 : M_{ts} \geq u) > \frac{\mathbb{E}[M_{00}]}{u}.$$

Some Intuition. Recall our earlier example: $\mathcal{K}_{ts} = \mathcal{K}_t^{(0)} + \mathcal{K}_s^{(1)}$. Then, for all $u > 0$,

$$\mathbb{P}(\exists t \geq 1 : \mathcal{K}_t^{(0)} \geq u) \leq \frac{\mathbb{E}[\mathcal{K}_0^{(0)}]}{u}, \quad \text{and} \quad \mathbb{P}(\exists s \geq 1 : \mathcal{K}_s^{(1)} \geq u) \leq \frac{\mathbb{E}[\mathcal{K}_0^{(1)}]}{u},$$

hence,

$$\mathbb{P}(\exists t, s \geq 1 : \mathcal{K}_{ts} \geq u) \leq \frac{2\mathbb{E}[\mathcal{K}_{00}]}{u}.$$

Analogues of Ville′s Inequality nevertheless hold

Definition (Cairoli and Walsh, 1970). $(\mathcal{F}_{ts})$ is said to be conditionally independent if

$$\mathcal{F}_{Ts} \perp\!\!\!\perp \mathcal{F}_{tS} \mid \mathcal{F}_{ts}, \quad \forall t \leq T, s \leq S.$$

Analogues of Ville′s Inequality nevertheless hold

Definition (Cairoli and Walsh, 1970). $(\mathcal{F}_{ts})$ is said to be conditionally independent if

$$\mathcal{F}_{Ts} \perp\!\!\!\perp \mathcal{F}_{tS} \mid \mathcal{F}_{ts}, \quad \forall t \leq T, s \leq S.$$

Intuition. Let (M_{ts}) be a partially ordered forward reverse martingale. By definition,

$$\mathbb{E}[M_{Ts}|\mathcal{F}_{ts}] = \mathbb{E}[M_{ts}|\mathcal{F}_{ts}].$$

Analogues of Ville′s Inequality nevertheless hold

Definition (Cairoli and Walsh, 1970). $(\mathcal{F}_{ts})$ is said to be conditionally independent if

$$\mathcal{F}_{Ts} \amalg \mathcal{F}_{tS} \mid \mathcal{F}_{ts}, \quad \forall t \leq T, s \leq S.$$

Intuition. Let (M_{ts}) be a partially ordered forward reverse martingale. By definition,

$$\mathbb{E}[M_{Ts}|\mathcal{F}_{ts}] = \mathbb{E}[M_{ts}|\mathcal{F}_{ts}].$$

Under conditional independence, one further has:

$$\mathbb{E}[M_{Ts}|\mathcal{F}_{tS}] = \mathbb{E}[M_{ts}|\mathcal{F}_{tS}].$$

Analogues of Ville's Inequality nevertheless hold

Theorem (Follows from Christofides and Serfling, 1990). Let (M_{ts}) be a partially ordered forward reverse martingale with respect to a conditionally independent filtration $(\mathcal{F}_{ts})$. Assume M_{ts} admits $k > 1$ moments. Then,

$$\mathbb{P}(\exists t, s \geq 1 : M_{ts} \geq u) \leq \left(\frac{k}{k-1}\right)^k \frac{\mathbb{E}[M_{00}^k]}{u^k}.$$

Analogues of Ville's Inequality nevertheless hold

Theorem (Follows from Christofides and Serfling, 1990). Let (M_{ts}) be a partially ordered forward reverse martingale with respect to a conditionally independent filtration $(\mathcal{F}_{ts})$. Assume M_{ts} admits $k > 1$ moments. Then,

$$\mathbb{P}(\exists t, s \geq 1 : M_{ts} \geq u) \leq \left(\frac{k}{k-1}\right)^k \frac{\mathbb{E}[M_{00}^k]}{u^k}.$$

Main Proof Idea. The following bound holds by Markov but is not useful:

$$\mathbb{P}(\exists t, s \geq 1 : M_{ts} \geq u) = \mathbb{P}\left(\max_{t,s\geq 1} M_{ts} \geq u\right) \leq \frac{1}{u^k}\mathbb{E}\left(\max_{t,s\geq 1} M_{ts}^k\right).$$

Analogues of Ville's Inequality nevertheless hold

Theorem (Follows from Christofides and Serfling, 1990). Let (M_{ts}) be a partially ordered forward reverse martingale with respect to a conditionally independent filtration $(\mathcal{F}_{ts})$. Assume M_{ts} admits $k > 1$ moments. Then,

$$\mathbb{P}(\exists t, s \geq 1 : M_{ts} \geq u) \leq \left(\frac{k}{k-1}\right)^k \frac{\mathbb{E}[M_{00}^k]}{u^k}.$$

Main Proof Idea. The following bound holds by Markov but is not useful:

$$\mathbb{P}(\exists t, s \geq 1 : M_{ts} \geq u) = \mathbb{P}\left(\max_{t,s\geq 1} M_{ts} \geq u\right) \leq \frac{1}{u^k}\mathbb{E}\left(\max_{t,s\geq 1} M_{ts}^k\right).$$

Instead, use conditional independence to show

$$\mathbb{P}(\exists t, s \geq 1 : M_{ts} \geq u) \leq \frac{1}{u^k}\mathbb{E}\left(\max_{t\geq 1} M_{t0}^k\right).$$

Analogues of Ville′s Inequality nevertheless hold

Theorem (Follows from Christofides and Serfling, 1990). Let (M_{ts}) be a partially ordered forward reverse martingale with respect to a conditionally independent filtration $(\mathcal{F}_{ts})$. Assume M_{ts} admits $k > 1$ moments. Then,

$$\mathbb{P}(\exists t, s \geq 1 : M_{ts} \geq u) \leq \left(\frac{k}{k-1} \right)^k \frac{\mathbb{E}[M_{00}^k]}{u^k}.$$

Proposition. If Ψ is convex, then $(\Psi(P_t, Q_s))_{t,s=1}^{\infty}$ is a partially-ordered reverse submartingale with respect to the filtration

$$\mathcal{E}_{ts} = \mathcal{E}_t^X \bigvee \mathcal{E}_s^Y := \sigma(\mathcal{E}_t^X \cup \mathcal{E}_s^Y), \quad t, s = 1, 2, \ldots,$$

where $(\mathcal{E}_t^X)$ and $(\mathcal{E}_s^Y)$ are the exchangeable filtrations generated by (X_t) and (Y_s) respectively. Furthermore, $(\mathcal{E}_{ts})$ is conditionally independent.

Application to convex divergences

Assume Ψ is convex and

$$\mathbb{E}\left[\exp\left(\lambda(\Psi(P_t, Q_s) - \Psi(P, Q))\right)\right] \leq \exp(\psi_{ts}(\lambda)),$$

- **Known Ordering.** Assume the orderings $T(n)$ and $S(n)$ are known and recall that $\sum_{k=1}^{\infty} 1/\ell(k) = 1$. Then,

$$\mathbb{P}\left\{\exists n \geq 1 : \Psi(P_{T(n)}, Q_{S(n)}) \geq \Psi(P, Q) + (\psi^*_{T(n)/2, S(n)/2})^{-1}\Big(\log \ell(\log_2 n) + \log(1/\delta)\Big)\right\} \leq \alpha.$$

Application to convex divergences

Assume Ψ is convex and

$$\mathbb{E}\Big[\exp\big(\lambda(\Psi(P_t, Q_s) - \Psi(P, Q))\big)\Big] \leq \exp(\psi_{ts}(\lambda)),$$

- **Known Ordering.** Assume the orderings $T(n)$ and $S(n)$ are known and recall that $\sum_{k=1}^{\infty} 1/\ell(k) = 1$. Then,

$$\mathbb{P}\left\{\exists n \geq 1 : \Psi(P_{T(n)}, Q_{S(n)}) \geq \Psi(P, Q) + (\psi^*_{T(n)/2, S(n)/2})^{-1}\Big(\log \ell(\log_2 n) + \log(1/\delta)\Big)\right\} \leq \alpha.$$

- **Partial Ordering.** : Assume now that

$$\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \frac{1}{g(k+j)} = \frac{1}{e}.$$

$$\mathbb{P}\left\{\exists t, s \geq 1 : \Psi(P_t, Q_s) \geq \Psi(P, Q) + (\psi^*_{t/2, s/2})^{-1}\Big(\log g(\log_2 t + \log_2 s) + \log(1/\delta)\Big)\right\} \leq \alpha.$$

Example: Maximum Mean Discrepancy, again

- **Known Ordering.** Assume the orderings $T(n)$ and $S(n)$ are known or depend only on external randomization. Then,

$$\mathbb{P}(\forall n \geq 1 : \ell_n \leq D(P\|Q) - D(P_{T(n)}\|Q_{S(n)}) \leq u_n) \geq 1 - \alpha.$$

$$\ell_n = 4\sqrt{\frac{B}{n}\Big[1.2\log\log n + \log(9/\alpha)\Big]} + 2\sqrt{\frac{2B}{n}}, \quad u_n = 2\sqrt{\frac{B}{n}\Big[1.2\log\log n + \log(9/\alpha)\Big]}.$$

Example: Maximum Mean Discrepancy, again

- **Known Ordering.** Assume the orderings $T(n)$ and $S(n)$ are known or depend only on external randomization. Then,

$$\mathbb{P}(\forall n \geq 1 : \ell_n \leq D(P\|Q) - D(P_{T(n)}\|Q_{S(n)}) \leq u_n) \geq 1 - \alpha.$$

$$\ell_n = 4\sqrt{\frac{B}{n}\left[1.2\log\log n + \log(9/\alpha)\right]} + 2\sqrt{\frac{2B}{n}}, \quad u_n = 2\sqrt{\frac{B}{n}\left[1.2\log\log n + \log(9/\alpha)\right]}.$$

- **Partial Ordering.** Instead,

$$\mathbb{P}(\forall t, s \geq 1 : \ell_{ts} \leq D(P\|Q) - D(P_n\|Q) \leq u_{ts}) \geq 1 - \alpha.$$

$$\ell_{ts} = 4\sqrt{\frac{Bts}{t+s}\left[2.2\log(\log t + \log s) + \log(15/\delta)\right]} + 2\sqrt{2B}\left(\frac{1}{\sqrt{t}} + \frac{1}{\sqrt{s}}\right)$$

$$u_{ts} = 2\sqrt{\frac{B}{t}\left[1.2\log\log t + \log(18/\alpha)\right]} + 2\sqrt{\frac{B}{s}\left[1.2\log\log s + \log(18/\alpha)\right]}.$$

Summary

- Reverse martingales can be useful tools for sequential inference.

Summary

- Reverse martingales can be useful tools for sequential inference.
- Forward and reverse martingales share many similar properties, but the latter arise more naturally from symmetry arguments.

Summary

- Reverse martingales can be useful tools for sequential inference.
- Forward and reverse martingales share many similar properties, but the latter arise more naturally from symmetry arguments.
- Filtrations are not nuisances—they are design tools.

Summary

- Reverse martingales can be useful tools for sequential inference.
- Forward and reverse martingales share many similar properties, but the latter arise more naturally from symmetry arguments.
- Filtrations are not nuisances—they are design tools.
- Partially-ordered martingales are **sometimes** useful for sequential inference, but have some technical challenges.

Summary

- Reverse martingales can be useful tools for sequential inference.
- Forward and reverse martingales share many similar properties, but the latter arise more naturally from symmetry arguments.
- Filtrations are not nuisances—they are design tools.
- Partially-ordered martingales are **sometimes** useful for sequential inference, but have some technical challenges.
- Let us know if you have a betting interpretation of reverse martingales. :)