Universal Off-Policy Evaluation

Xuhui Liu

Nanjing University

June 18, 2021

Yash Chandak et al.
"Universal Off-Policy Evaluation." CoRR, 2021.

Table of Contents

Table of Contents

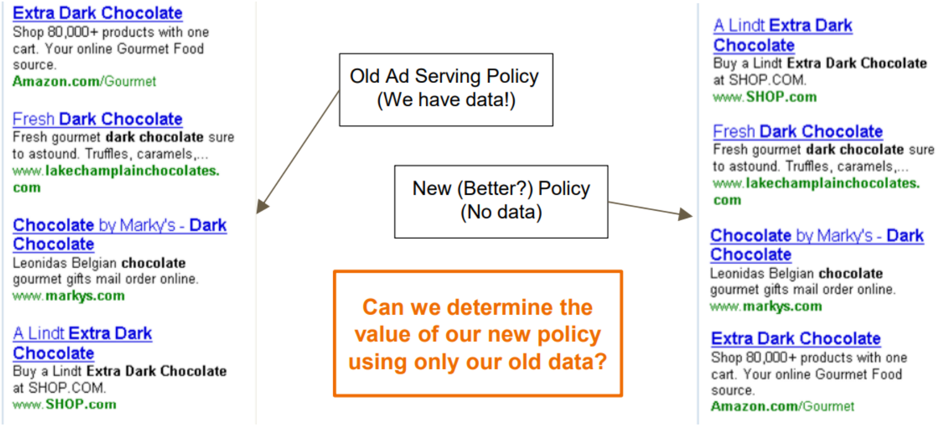
Off-Policy Evaluation

Definition: The problem of evaluating a **new strategy** for behavior, or policy, using only observations collected during the execution of **another policy**.

Motivation

- Want to evaluate new method without incurring the risk and cost of actually implementing this new method/policy.
- Existing logs containing huge amounts of historical data based on existing policies.
- It makes economical sense to, if possible, use these logs.
- It makes economical sense to, if possible, not risk the loss of testing out a new potentially bad policy.
- Online ad placement is a good example.

Motivation



Extra Dark Chocolate
Shop 80,000+ products with one cart. Your online Gourmet Food source.
Amazon.com/Gourmet
Fresh Dark Chocolate
Fresh gourmet dark chocolate sure to astound. Truffles, caramels,...
www.lakechamplainchocolates.com
Chocolate by Marky's - Dark Chocolate
Leonidas Belgian chocolate gourmet gifts mail order online.
www.markys.com
A Lindt Extra Dark Chocolate
Buy a Lindt Extra Dark Chocolate at SHOP.COM.
www.SHOP.com
Old Ad Serving Policy (We have data!)
New (Better?) Policy (No data)
Can we determine the value of our new policy using only our old data?
A Lindt Extra Dark Chocolate
Buy a Lindt Extra Dark Chocolate at SHOP.COM.
www.SHOP.com
Fresh Dark Chocolate
Fresh gourmet dark chocolate sure to astound. Truffles, caramels,...
www.lakechamplainchocolates.com
Chocolate by Marky's - Dark Chocolate
Leonidas Belgian chocolate gourmet gifts mail order online.
www.markys.com
Extra Dark Chocolate
Shop 80,000+ products with one cart. Your online Gourmet Food source.
Amazon.com/Gourmet

Importance Sampling

- Naive importance sampling

$$
\begin{aligned}
J\left(\pi_\theta\right) &= \mathbb{E}_{\tau \sim \pi_\beta(\tau)}\left[\frac{\pi_\theta(\tau)}{\pi_\beta(\tau)} \sum_{t=0}^{H} \gamma^t r(\mathbf{s}, \mathbf{a})\right] \\
&= \mathbb{E}_{\tau \sim \pi_\beta(\tau)}\left[\left(\prod_{t=0}^{H} \frac{\pi_\theta\left(\mathbf{a}_t \mid \mathbf{s}_t\right)}{\pi_\beta\left(\mathbf{a}_t \mid \mathbf{s}_t\right)}\right) \sum_{t=0}^{H} \gamma^t r(\mathbf{s}, \mathbf{a})\right] \approx \sum_{i=1}^{n} w_H^i \sum_{t=0}^{H} \gamma^t r_t^i \\
&\text{where } w_t^i = \frac{1}{n} \Pi_{t'=0}^{t} \frac{\pi_\theta\left(a_{t'}^i \mid s_{t'}^i\right)}{\pi_\beta\left(a_{t'}^i \mid s_{t'}^i\right)}
\end{aligned}
$$

- Weighted importance sampling

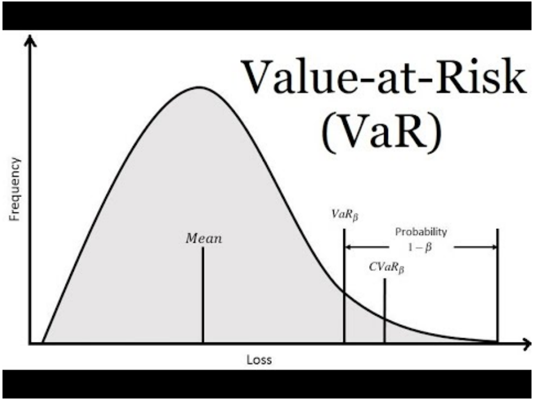
$$
w_t^i = \frac{1}{n} \Pi_{t'=0}^{t} \frac{\pi_\theta\left(a_{t'}^i \mid s_{t'}^i\right)}{\pi_\beta\left(a_{t'}^i \mid s_{t'}^i\right)} \quad \Rightarrow \quad w_t^i = \frac{1}{\sum_{i=1}^{n} w_t^i} \Pi_{t'=0}^{t} \frac{\pi_\theta\left(a_{t'}^i \mid s_{t'}^i\right)}{\pi_\beta\left(a_{t'}^i \mid s_{t'}^i\right)}
$$

Statistical Quantities

Consider a random variable X, the cumulative distribution function and probability distribution function of which are $F(x)$ and $p(x)$, respectively.

- Mean $\mathbb{E}[X]$.
- Quantile $\mathrm{quantile}_\alpha(X) = F^{-1}(\alpha)$.
- Value at Risk (VaR) $\mathrm{VaR}_\alpha(X) = \mathrm{quantile}_\alpha$.
- Conditional Value at Risk (CVaR) $\mathrm{CVaR}_\alpha(X) = \mathbb{E}[X | X \leq \mathrm{quantile}_\alpha(X)]$.
- Variance $\mathbb{E}[(X - \mathbb{E}X)^2]$.
- Entropy $H(X) = \int p(x) \log p(x) dx$.

Illustration



Limitations

- Safety critical applications, e.g., automated healthcare.
  Risk-prone metrics: **Value at risk (VaR)** and **conditional value at risk (CVaR)**.

- Applications like online recommendations are subject to noisy data
  Robust metrics: **Median** and **other quantiles**.

- Applications involving direct human-machine interaction, such as robotics and autonomous driving.
  Uncertainty metrics: **Variance** and **entropy**.

How do we develop a universal off-policy method—one that can estimate **any desired performance metrics** and can also provide **high-confidence bounds** that hold simultaneously with high probability for those metrics?

Table of Contents

Notations

- Partial Observable Markov Decision Process (POMDP) $(\mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{P}, \Omega, \mathcal{R}, \gamma, d_0)$.
- $\mathcal{D}$ is the data set $(H_i)_{i=1}^n$ collected using behavior policies $(\beta_i)_{i=1}^n$, where H_i is the observed history $(O_0, A_0, \beta(A_0|O_0), R_0, O_1, \dots)$.
- $G_i := \sum_{j=0}^T \gamma^j R_j$ is the return for H_i.
- T is the horizon length.
- G_π and H_π is the random variables for return s and comlete trajectories under any policy π, respectively.
- $g(h)$ is the return for trajectory h.
- $\mathcal{H}_\pi$ is the set of all possible trajectories for policy π.

Assumptions

Assumption 1

The set $\mathcal{D}$ contains independent (not necessarily identically distributed) histories generated using $(\beta_i)_{i=1}^{n}$, along with the probability of the actions chosen, such that for some $\epsilon > 0$, $(\beta_i(a|o) < \epsilon) \implies (\pi(a|o) = 0)$, for all $o \in \mathcal{O}$, $a \in \mathcal{A}$, and $i \in (1, \ldots, n)$.

Method

- First estimate its cumulative distribution F_π.
- Then use it to estimate its parameter $\psi(F_\pi)$.

Method

- Represent F_π with return G_π.

$$F_\pi(\nu) = \Pr\left(G_\pi \leq \nu\right) = \sum_{x \in \mathcal{X}, x \leq \nu} \Pr\left(G_\pi = x\right) = \sum_{x \in \mathcal{X}, x \leq \nu} \left(\sum_{h \in \mathcal{H}_\pi} \Pr\left(H_\pi = h\right) \mathbb{1}_{\{g(h) = x\}} \right)$$

- Exchange the order of Sum.

$$F_\pi(\nu) = \sum_{h \in \mathcal{H}_\pi} \Pr\left(H_\pi = h\right) \sum_{x \in \mathcal{X}, x \leq \nu} \mathbb{1}_{\{g(h) = x\}} = \sum_{h \in \mathcal{H}_\pi} \Pr\left(H_\pi = h\right) \left(\mathbb{1}_{\{g(h) \leq \nu\}} \right)$$

Method

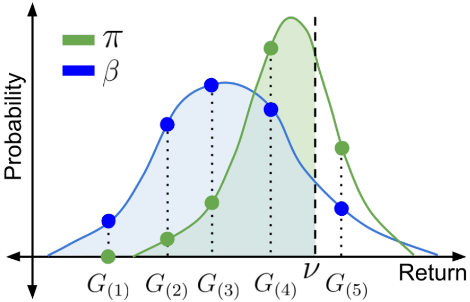
- From Assumption 1 $\forall \beta, \mathcal{H}_\pi \subset \mathcal{H}_\beta$,

$$F_\pi(\nu) = \sum_{h \in \mathcal{H}_\beta} \Pr\left(H_\pi = h\right)\left(\mathbb{1}_{\{g(h) \leq \nu\}}\right) = \sum_{h \in \mathcal{H}_\beta} \Pr\left(H_\beta = h\right) \frac{\Pr\left(H_\pi = h\right)}{\Pr\left(H_\beta = h\right)} \left(\mathbb{1}_{\{g(h) \leq \nu\}}\right)$$

- Let $\rho_i := \Pi_{j=0}^{T} \frac{\pi(A_j|O_j)}{\beta_i(A_j|O_j)}$, which is equal to $\Pr(H_\pi = h)/\Pr(H_\beta = h)$.

$$\forall \nu \in \mathbb{R}, \quad \hat{F}_n(\nu) := \frac{1}{n} \sum_{i=1}^{n} \rho_i \mathbb{1}_{\{G_i \leq \nu\}}$$

Illustration



Partial Observable Setting

- $\mathcal{O}, \tilde{\mathcal{O}}$: Observation set for the behavior policy and the evaluation policy.
- If $\tilde{\mathcal{O}} = \mathcal{O} = \mathcal{S}$, it becomes MDP setting.
- If $\tilde{\mathcal{O}} = \mathcal{O}$, as $\beta(a|o) = \beta(a|\tilde{o})$, one can use density estimation on the available data, $\mathcal{D}$, to construct an estimator $\hat{\beta}(a|o)$ of $\Pr(a|\tilde{o}) = \beta(a|\tilde{o})$.
- $\tilde{\mathcal{O}} \neq \mathcal{O}$, it is only possible to estimate $\Pr(a|\tilde{o}) = \sum_{x \in o} \beta(a|x) \Pr(x|\tilde{o})$ through density estimation using data $\mathcal{D}$.

Probability Distribution and Inverse CDF

- Let $(G_{(i)})_{i=1}^{n}$ be the order statistics for samples $(G_i)_{i=1}^{n}$ and $G_0 := G_{\min}$.
- Inverse CDF
  $\hat{F}_n^{-1}(\alpha) := \min\left\{g \in \left(G_{(i)}\right)_{i=1}^{n} \mid \hat{F}_n(g) \geq \alpha\right\}$
- Probability distribution
  $\mathrm{d}\hat{F}_n\left(G_{(i)}\right) := \hat{F}_n\left(G_{(i)}\right) - \hat{F}_n\left(G_{(i-1)}\right)$

Estimator

- $\mu_\pi\left(\hat{F}_n\right) := \sum_{i=1}^n \, \mathrm{d}\hat{F}_n\left(G_{(i)}\right) G_{(i)}$,
- $\sigma_\pi^2\left(\hat{F}_n\right) := \sum_{i=1}^n \, \mathrm{d}\hat{F}_n\left(G_{(i)}\right) \left(G_{(i)} - \mu_\pi\left(\hat{F}_n\right)\right)^2$,
- $Q_\pi^\alpha\left(\hat{F}_n\right) := \hat{F}_n^{-1}(\alpha)$,
- $\mathrm{CVaR}_\pi^\alpha\left(\hat{F}_n\right) := \frac{1}{\alpha} \sum_{i=1}^n \, \mathrm{d}\hat{F}_n\left(G_{(i)}\right) G_{(i)} \mathbb{1}_{\left\{G_{(i)} \leq Q_\pi^\alpha\left(\hat{F}_n\right)\right\}}$.
  - Definition of CVaR:
  $$\mathrm{CVaR}_\pi^\alpha\left(F_\pi\right) = \mathbb{E}\left[G_\pi \mid G_\pi \leq F_\pi^{-1}(\alpha)\right]$$

Table of Contents

High-Confidence Bounds

1. It's easy to obtain bounds for a single point.
2. It's hard hard to obtain bounds for a interval.
3. CDF is monotonically non-decreasing.

- Let $(\kappa_i)_{i=1}^{K}$ be any K "key points at which we obtain confidence interval for $(F_\pi(\kappa_i))_{i=1}^{K}$.
- Generalize to whole interval based on these "key points".

Let $\mathrm{CI}_{\_}(\kappa_i, \delta_i)$ and $\mathrm{CI}_{+}(\kappa_i, \delta_i)$ be the lower and the upper confidence bounds on $F_\pi(\kappa_i)$, respectively, such that

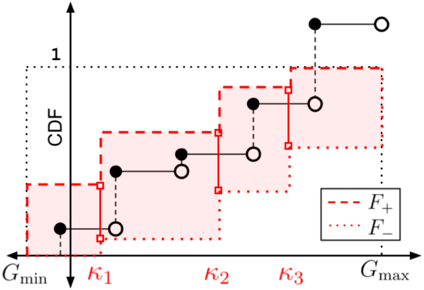
$$\forall i \in (1, \dots, K), \quad \Pr\left(\mathrm{CI}_{\_}(\kappa_i, \delta_i) \leq F_\pi(\kappa_i) \leq \mathrm{CI}_{+}(\kappa_i, \delta_i)\right) \geq 1 - \delta_i$$

Based on this constuction, we formulate a lower function $F_{\_}$ and an upper function F_{+}:

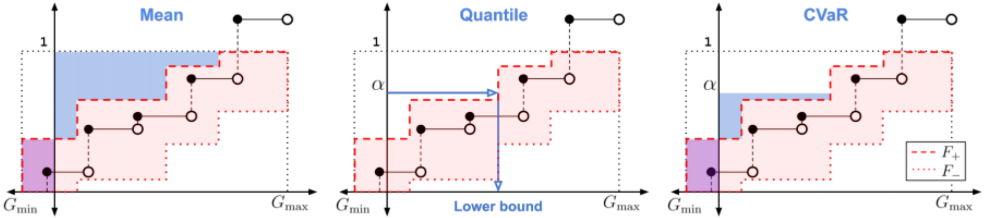
$$F_{\_}(\nu) := \begin{cases} 1 & \text{if } \nu > G_{\max} \\ \max_{\kappa_i \leq \nu} CI_{\_}(\kappa_i, \delta_i) & \text{otherwise} \end{cases}$$

$$F_{+}(\nu) := \begin{cases} 0 & \text{if } \nu < G_{\min} \\ \min_{\kappa_i \geq \nu} CI_{+}(\kappa_i, \delta_i) & \text{otherwise} \end{cases}$$

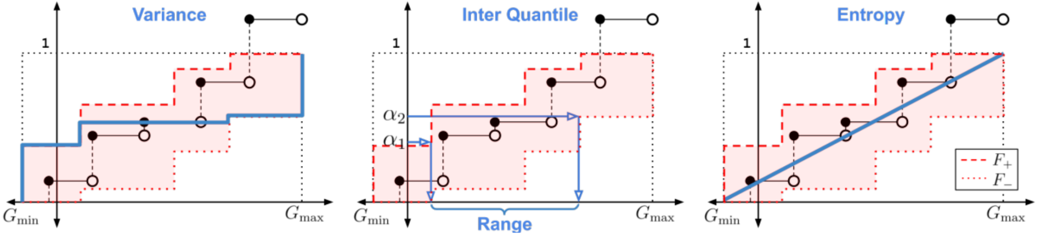
Illustration



Illustration



Illustration



Bootstrap Bounds[1]

Algorithm 1: Bootstrap Bounds for $\psi(F_\pi)$

1. **Input:** Dataset $\mathcal{D}$, Confidence level $1-\delta$
2. Bootstrap B datasets $(\mathcal{D}_i^*)_{i=1}^B$ and create $(\bar{F}_{n,i}^*)_{i=1}^B$
3. Bootstrap estimates $(\psi(\bar{F}_{n,i}^*))_{i=1}^B$ using $(\bar{F}_{n,i}^*)_{i=1}^B$
4. Compute (ψ_-, ψ_+) using BCa$((\psi(\bar{F}_{n,i}^*))_{i=1}^B, \delta)$
5. **Return** (ψ_-, ψ_+)

[1] B. Efron and R. J. Tibshirani. "An introduction to the Bootstrap". CRC press, 1994.

Non-stationary

Assumption 3

For any $\nu, \exists w_\nu \in \mathbb{R}^d$, such that, $\forall i \in [1, L+\ell], \quad F_\pi^{(i)}(\nu) = \phi(i)^\top w_\nu$.

- Estimating $F_\pi^{(i)}$ can now be seen as a time-series forecasting problem.
- Wild bootstrap[2] provides approximate CIs.

---

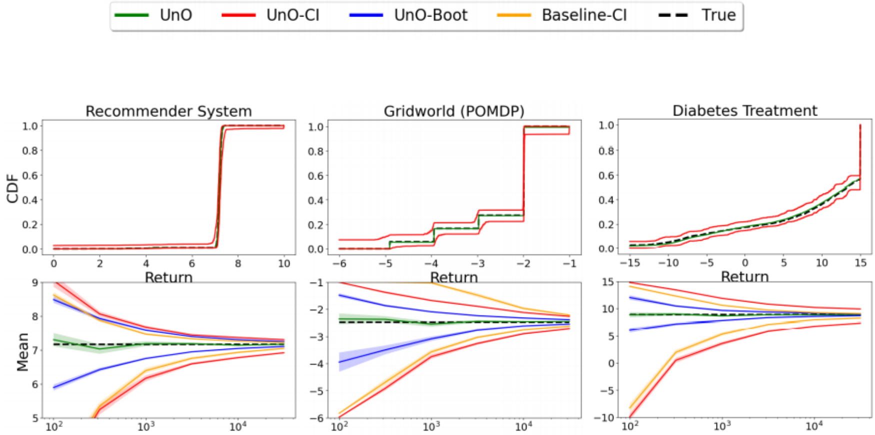
[2] E. Mammen. "Bootstrap and wild bootstrap for high dimensional linear models." The Annals of Statistics, pages 255–285, 1993.

Table of Contents

Setting

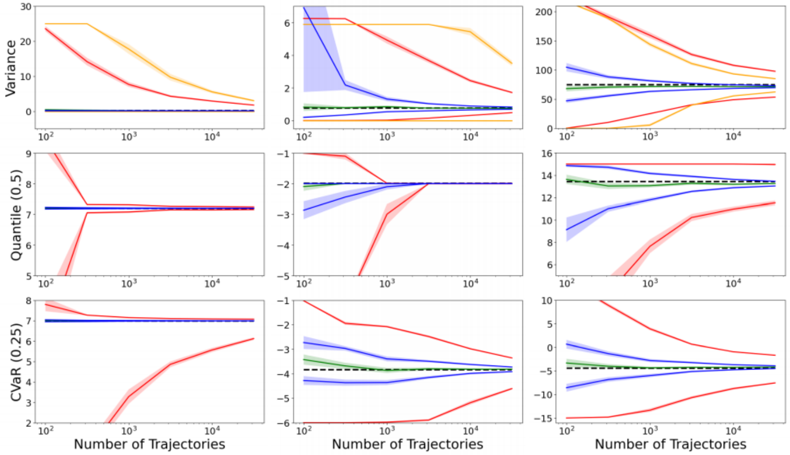
- A simulated stationary and a non-stationary **recommender system domain**, where the user's interest for a finite set of items is represented using the corresponding item's reward.
- **Type-1 Diabetes Mellitus Simulator (T1DMS)** for the treatment of type-1 diabetes.
- A **continuous-state Gridworld with partial observability** (which also makes the domain non-Markovian in the observations), stochastic transitions, and eight discrete actions corresponding to up, down, left, right, and the four diagonal movements.

Experiments[3]



[3] P. Thomas, G. Theocharous, and M. Ghavamzadeh. "High confidence policy improvement." ICML, 2015.

Experiments[4]



[4] Y. Chandak, S. Shankar, and P. S. Thomas. "High confidence off-policy (or counterfactual) variance estimation." AAAI, 2021.

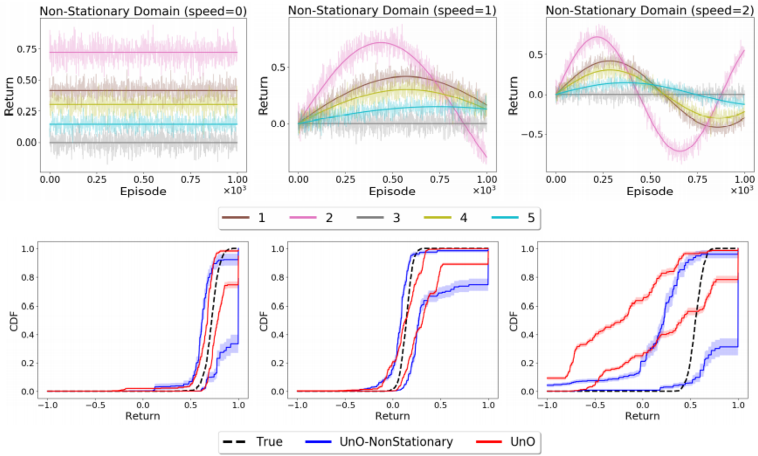

Non-Stationary Domain (speed=0)
Non-Stationary Domain (speed=1)
Non-Stationary Domain (speed=2)
Return
Episode
×10³
1
2
3
4
5
CDF
Return
True
UnO-NonStationary
UnO

Table of Contents

Theorem

- Estimator

$$\forall \nu \in \mathbb{R}, \quad \hat{F}_n(\nu) := \frac{1}{n} \sum_{i=1}^{n} \rho_i \mathbb{1}_{\{G_i \leq \nu\}}$$

- Theoretical guarantee

Theorem 1. *Under Assumption 1, $\hat{F}_n$ is an unbiased and uniformly consistent estimator of F_π,*

$$\forall \nu \in \mathbb{R}, \quad \mathbb{E}_{\mathcal{D}}\left[\hat{F}_n(\nu)\right] = F_\pi, \qquad \sup_{\nu \in \mathbb{R}} \left|\hat{F}_n(\nu) - F_\pi(\nu)\right| \xrightarrow{a.s.} 0.$$

Part 1 Unbiasedness

Recall that

$$F_\pi(\nu) = \sum_{h \in \mathcal{H}_\beta} \Pr\left(H_\pi = h\right)\left(\mathbb{1}_{\{g(h) \leq \nu\}}\right) = \sum_{h \in \mathcal{H}_\beta} \Pr\left(H_\beta = h\right) \frac{\Pr\left(H_\pi = h\right)}{\Pr\left(H_\beta = h\right)}\left(\mathbb{1}_{\{g(h) \leq \nu\}}\right)$$

The probability of a trajectory under a policy π with partial observations and non-Markovian structure is

$$\begin{aligned}\Pr\left(H_\pi = h\right) &= \Pr\left(s_0\right)\Pr\left(o_0 \mid s_0\right)\Pr\left(\tilde{o}_0 \mid o_0, s_0\right)\Pr\left(a_0 \mid s_0, o_0, \tilde{o}_0; \pi\right) \\ &\times \prod_{i=0}^{T-1}\left(\Pr\left(r_i \mid h_i\right)\Pr\left(s_{i+1} \mid h_i\right)\Pr\left(o_{i+1} \mid s_{i+1}, h_i\right)\Pr\left(\tilde{o}_{i+1} \mid s_{i+1}, o_{i+1}, h_i\right)\right. \\ &\quad \left.\times \Pr\left(a_{i+1} \mid s_{i+1}, o_{i+1}, \tilde{o}_{i+1}, h_i; \pi\right)\right)\Pr\left(r_T \mid h_T\right)\end{aligned}$$

The ratio can be written as

$$\begin{aligned}\frac{\Pr\left(H_\pi = h\right)}{\Pr\left(H_\beta = h\right)} &= \frac{\Pr\left(a_0 \mid s_0, o_0, \tilde{o}_0; \pi\right)}{\Pr\left(a_0 \mid s_0, o_0, \tilde{o}_0; \beta\right)} \prod_{i=0}^{T-1} \frac{\Pr\left(a_{i+1} \mid s_{i+1}, o_{i+1}, \tilde{o}_{i+1}, h_i; \pi\right)}{\Pr\left(a_{i+1} \mid s_{i+1}, o_{i+1}, \tilde{o}_{i+1}, h_i; \beta\right)} \\ &= \prod_{i=0}^{T} \frac{\pi\left(a_i \mid \tilde{o}_i\right)}{\beta\left(a_i \mid o_i\right)} \\ &= \rho(h)\end{aligned}$$

Then we have

$$F_\pi(\nu) = \sum_{h \in \mathcal{H}_\beta} \Pr(H_\beta = h)\rho(h)(\mathbb{1}_{\{g(h) \leq \nu\}}).$$

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}}\left[\hat{F}_n(\nu)\right] &= \mathbb{E}_{\mathcal{D}}\left[\frac{1}{n}\sum_{i=1}^{n}\rho_i\left(\mathbb{1}_{\{G_i\leq\nu\}}\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{D}}\left[\rho_i\left(\mathbb{1}_{\{G_i\leq\nu\}}\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{h\in\mathcal{H}_{\beta_i}}\Pr\left(H_{\beta_i}=h\right)\rho(h)\left(\mathbb{1}_{\{g(h)\leq\nu\}}\right) \\
&\overset{(a)}{=} \frac{1}{n}\sum_{i=1}^{n}F_\pi(\nu) \\
&= F_\pi(\nu)
\end{aligned}
$$

Part 2 Uniform Consistency

- First we show pointwise consistency, i.e., for all ν, $\hat{F}_n(\nu) \xrightarrow{\text{a.s.}} F_\pi(\nu)$.
- Then we use this to establish uniform consistency.

Kolmogorov' Strong Law of Large Numbers

The strong law of large numbers states that the sample average converges almost surely to the expected value:

$$\bar{X}_n \xrightarrow{\text{a.s.}} \mu \quad \text{when } n \to \infty$$

if one of the following conditions is satisfied:

1. The random variables are identically distributed;
2. For each n, the variance of X_n is finite, and

$$\sum_{n=1}^{\infty} \frac{\operatorname{Var}[X_n]}{n^2} < \infty$$

- Let
$$X_i := \rho_i \left(\mathbb{1}_{\{G_i \leq \nu\}} \right)$$
- By assumption 1, $\beta(a|o) \geq \epsilon$ when $\pi(a|\tilde{o}) > 0$. This implies the ratio is bounded above, and hence X_i are bounded above and have a finite variance.
- By Kolmogorov's strong law of large numbers:

$$\hat{F}_n(\nu) = \frac{1}{n} \sum_{i=1}^{n} X_i \overset{\text{a.s.}}{\to} \mathbb{E}_{\mathcal{D}} \left[\frac{1}{n} \sum_{i=1}^{n} X_i \right] = F_\pi(\nu)$$

- Some extra notation to tackle discontinuities in CDF F_π

$$F_\pi\left(\nu^-\right) := \Pr\left(G_\pi < \nu\right) = F_\pi(\nu) - \Pr\left(F_\pi = \nu\right), \quad \hat{F}_n\left(\nu^-\right) := \frac{1}{n}\sum_{i=1}^{n} \rho_i \left(\mathbb{1}_{\{G_i<\nu\}}\right)$$

- Similarly, we have

$$\hat{F}_n\left(\nu^-\right) \xrightarrow{\text{a.s}} F_\pi\left(\nu^-\right)$$

Let $\epsilon_1 > 0$, and let K be any value more than $1/\epsilon_1$. Let $(\kappa_i)_{i=0}^{K}$ be K key points,

$$G_{\min} = \kappa_0 < \kappa_1 \leq \kappa_2 \ldots \leq \kappa_{K-1} < \kappa_K = G_{\max}$$

which create K intervals such that for all $i \in (1, \ldots, K-1)$,

$$F_\pi(\kappa_i^-) \leq \frac{i}{K} \leq F_\pi(\kappa_i)$$

Then by construction, if $\kappa_{i-1} < \kappa_i$,

$$F_\pi(\kappa_i^-) - F_\pi(\kappa_{i-1}) \leq \frac{i}{K} - \frac{i-1}{K} = \frac{1}{K} < \epsilon_1.$$

For any ν, let κ_{i-1} and κ_i be such that $\kappa_{i-1} \leq \nu < \kappa_i$. Then,

$$\begin{aligned}
\hat{F}_n(\nu) - F_\pi(\nu) &\leq \hat{F}_n\left(\kappa_i^-\right) - F_\pi\left(\kappa_{i-1}\right) \\
&\leq \hat{F}_n\left(\kappa_i^-\right) - F_\pi\left(\kappa_i^-\right) + \epsilon_1.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\hat{F}_n(\nu) - F_\pi(\nu) &\geq \hat{F}_n\left(\kappa_{i-1}\right) - F_\pi\left(\kappa_i^-\right) \\
&\geq \hat{F}_n\left(\kappa_{i-1}\right) - F_\pi\left(\kappa_{i-1}\right) - \epsilon_1
\end{aligned}$$

Then, $\forall \nu \in \mathbb{R}$,

$$\hat{F}_n(\kappa_{i-1}) - F_\pi(\kappa_{i-1}) - \epsilon_1 \le \hat{F}_n(\nu) - F_\pi(\nu) \le \hat{F}_n(\kappa_i^-) - F_\pi(\kappa_i^-) + \epsilon_1,$$

Let

$$\Delta_n := \max_{i \in (1 \ldots K-1)} \left\{ \left| \hat{F}_n(\kappa_i) - F_\pi(\kappa_i) \right|, \left| \hat{F}_n(\kappa_i^-) - F_\pi(\kappa_i^-) \right| \right\}$$

By the pointwise convergence, we have

$$\Delta_n \overset{\text{a.s.}}{\to} 0$$

and thus,

$$\left| \hat{F}_n(\nu) - F_\pi(\nu) \right| \le \Delta_n + \epsilon_1$$

Finally, since the inequality holds for $\forall \nu \in \mathbb{R}$ and is valid for any $\epsilon_1 > 0$, making $\epsilon_1 \to 0$ gives the desired result,

$$\sup_{\nu \in \mathbb{R}} \left| \hat{F}_n(\nu) - F_\pi(\nu) \right| \xrightarrow{\text{a.s.}} 0$$

Variance Reduction

- Inspired by weighted importance sampling

$$\forall \nu \in \mathbb{R}, \quad \bar{F}_n(\nu) := \frac{1}{\sum_{j=1}^{n} \rho_j} \left(\sum_{i=1}^{n} \rho_i \left(\mathbb{1}_{\{G_i \leq \nu\}} \right) \right).$$

- Under Assumption 1, $\bar{F}_n$ may be biased but is a uniformly consistent estimator of F_π,

$$\forall \nu \in \mathbb{R}, \quad \mathbb{E}_{\mathcal{D}} \left[\bar{F}_n(\nu) \right] \neq F_\pi, \quad \sup_{\nu \in \mathbb{R}} \left| \bar{F}_n(\nu) - F_\pi(\nu) \right| \xrightarrow{\text{a.s.}} 0$$

Part 1 Biased: We prove this using a counter-example. Let $n = 1$, so

$$\begin{aligned}
\forall \nu \in \mathbb{R}, \quad \mathbb{E}_{\mathcal{D}}\left[\bar{F}_n(\nu)\right] &= \mathbb{E}_{\mathcal{D}}\left[\frac{1}{\sum_{j=1}^{1}\rho_j}\left(\sum_{i=1}^{1}\rho_i \mathbb{1}_{\{G_i \leq \nu\}}\right)\right] \\
&= \mathbb{E}_{\mathcal{D}}\left[\mathbb{1}_{\{G_1 \leq \nu\}}\right] \\
&\overset{(a)}{=} \sum_{h \in \mathcal{H}_{\beta_1}} \Pr\left(H_{\beta_1} = h\right)\left(\mathbb{1}_{\{g(h) \leq \nu\}}\right) \\
&= F_{\beta_1}(\nu) \\
&\neq F_{\pi}(\nu)
\end{aligned}$$

Part 2 Uniform Consistency: First, we will establish pointwise consistency, i.e., for any $\nu, \bar{F}_n(\nu) \overset{\text{a.s.}}{\to} F_\pi(\nu)$, and then we will use this to establish uniform consistency, as required.

$$\forall \nu \in \mathbb{R}, \quad \bar{F}_n(\nu) = \frac{1}{\sum_{j=1}^{1} \rho_j} \left(\sum_{i=1}^{1} \rho_i \mathbb{1}_{\{G_i \leq \nu\}} \right)$$
$$= \left(\frac{1}{n} \sum_{j=1}^{n} \rho_j \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \rho_i \mathbb{1}_{\{G_i \leq \nu\}} \right).$$

If both $\left(\lim_{n \to T} \frac{1}{n} \sum_{j=1}^{n} \rho_j\right)^{-1}$ and $\left(\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \rho_i \mathbb{1}_{\{G_i \leq \nu\}}\right)$ exist, then using Slutsky's theorem, $\forall \nu \in \mathbb{R}$,

$$\lim_{n \to \infty} \bar{F}_n(\nu) = \left(\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \rho_j \right)^{-1} \left(\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \rho_i \mathbb{1}_{\{G_i \leq \nu\}} \right)$$

Notice using Kolmogorov's strong law of large numbers that the term in the first parentheses will converge to the expected value of importance ratios, which equals one. Similarly, we know that the term in the second parentheses will converge to $F_\pi(\nu)$ almost surely. Therefore,

$$\forall \nu \in \mathbb{R}, \quad \bar{F}_n(\nu) \xrightarrow{\text{a.s.}} (1)^{-1} \left(F_\pi(\nu)\right) = F_\pi(\nu)$$