Uncertainty-Aware Reinforcement Learning for Risk-Sensitive Player Evaluation in Sports Game

Guiliang Liu, Yudong Luo, Oliver Schulte, Pascal Poupart

University of Waterloo, The Chinese University of Hong Kong, Shenzhen, Vector Institute, and Simon Fraser University

Neurips 2022 Presentation

Problem Definition

Player Evaluation:

- **Definition:** Evaluate the contribution of players in the game (drafting, coaching, trading)
- Access to a dataset, e.g., game recordings
- Mainstream method: quantify player's action impact

- **Example:**

1) Supervised Learning: give a label of 1 to scoring a goal, predict the scoring probability of other actions
2) Reinforcement Learning: naturally has an action-value function, named Q value. Design the reward as 1 to scoring a goal, 0 otherwise

Problem Definition

Player Evaluation:

- **Definition:** Evaluate the contribution of players in the game (drafting, coaching, trading)
- Access to a dataset, e.g., game recordings
- Mainstream method: quantify action impacts.

- **Challenges:**

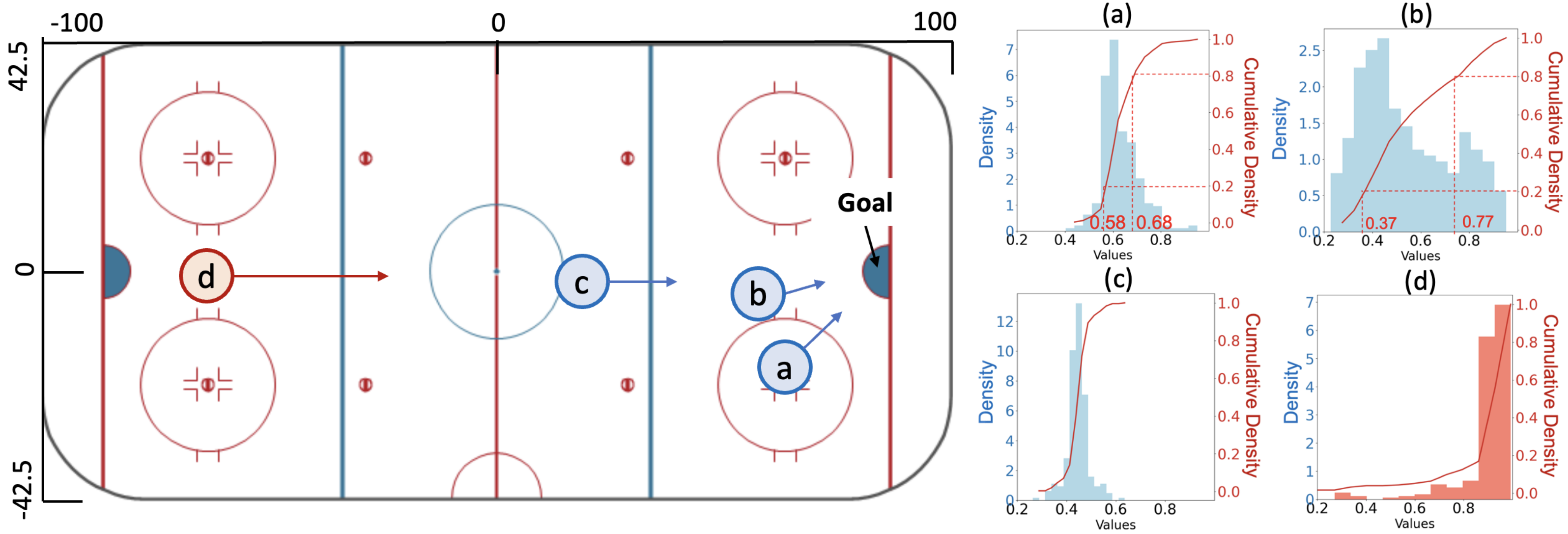
1) Previous methods are expectation-based, which cannot differentiate the risk-seeking actions from the risk-averse actions.
2) How to distinguish these actions and assign proper credits to the players remains a fundamental challenge in sports analytics.

- **Our solution:** Risk-Sensitive Player Evaluation with Post-hoc Calibration

Motivation

Example: The predicted distribution of future goals for the shots made at positions (a to d).

- **Risk-Sensitive Evaluation**: Distributions (a) and (b) have the same expectation (around 0.6). The first shot has a larger risk-averse estimate and a smaller risk-seeking estimate.
- **Post-hoc Calibration**: shot made at the position (d) is rare in an ice hockey game, and thus this event is likely to be OoD, leading to a biased prediction.



Uncertainty-Aware Reinforcement Learning

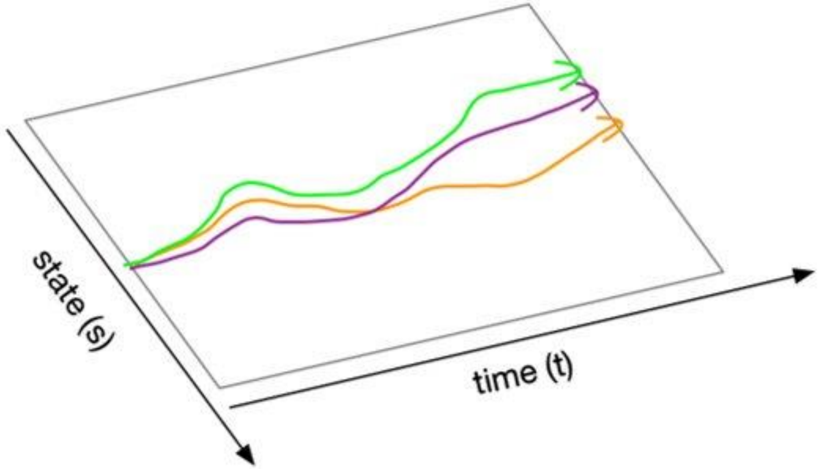
Source of uncertainties:

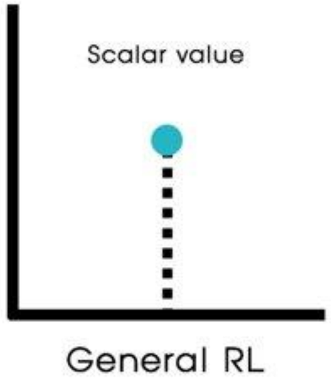
- **Aleatoric uncertainty:** the intrinsic uncertainty of the environment (sports game is highly stochastic)
- **Epistemic uncertainty:** due to lack of knowledge, e.g., limited date samples (we only have access to a dataset)

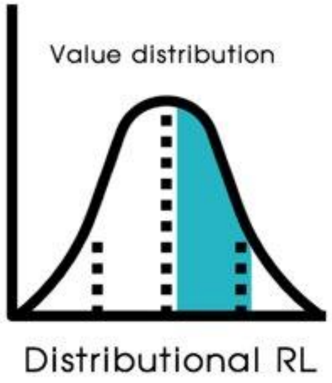
- In sports evaluation, we need to consider both uncertainties

Aleatoric Uncertainty: Distributional RL

Intrinsic uncertainty leads to value distribution:







- **Traditional RL:** only learn the mean value of the value distribution
- **Distributional RL:** learn the full value distribution

- Distributional Bellman Operator

$$\mathcal{T}^{\pi} Z_k(s_t, a_t) \triangleq R_k(s_t, a_t) + \gamma Z_k(S_{t+1}, A_{t+1})$$

Where $s_{t+1} \sim P_{\mathcal{T}}(S_{t+1}|s_t, a_t)$ and $a_{t+1} \sim \pi(A_{t+1}|S_{t+1})$

Aleatoric Uncertainty: Distributional RL

Intrinsic uncertainty leads to value distribution:

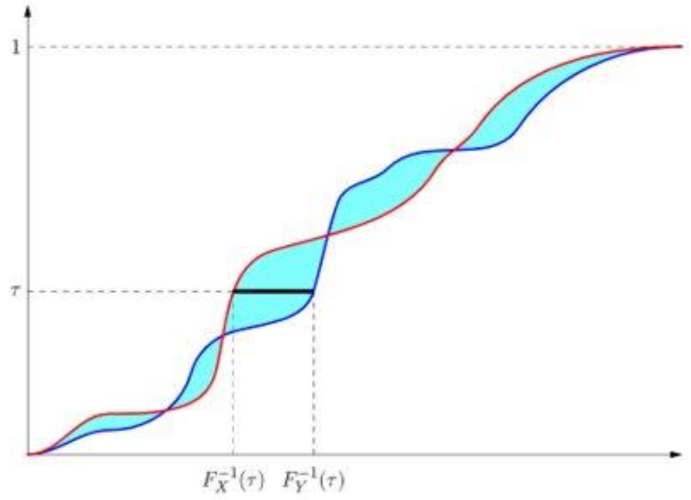
- Distributional Bellman Operator

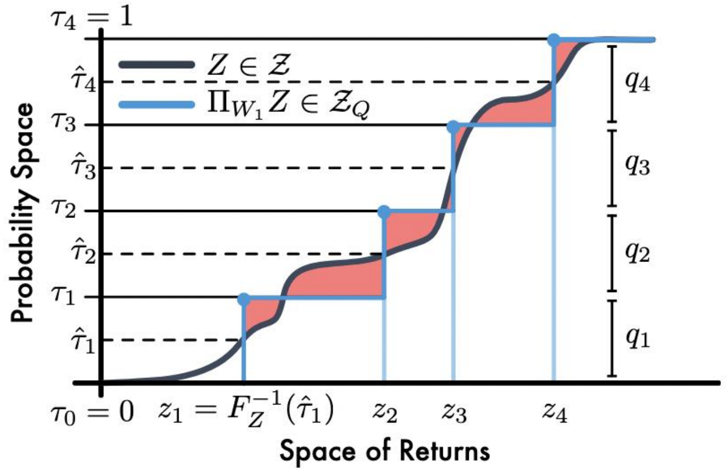
$$\mathcal{T}^{\pi} Z_k(s_t, a_t) \triangleq R_k(s_t, a_t) + \gamma Z_k(S_{t+1}, A_{t+1})$$

Where $s_{t+1} \sim P_{\mathcal{T}}(S_{t+1} | s_t, a_t)$ and $a_{t+1} \sim \pi(A_{t+1} | S_{t+1})$

- Converge? Distributional Bellman Operator is a contraction mapping under p-Wasserstein metric

$$W_p(U, Y) = \left(\int_0^1 |F_Y^{-1}(\omega) - F_U^{-1}(\omega)|^p d\omega \right)^{1/p}$$





- A lot of existing methods use quantile regression, representing quantile function by a mixture of N Diracs

Aleatoric Uncertainty: Distributional RL

- Distributional RL for **Aleatoric Uncertainty**

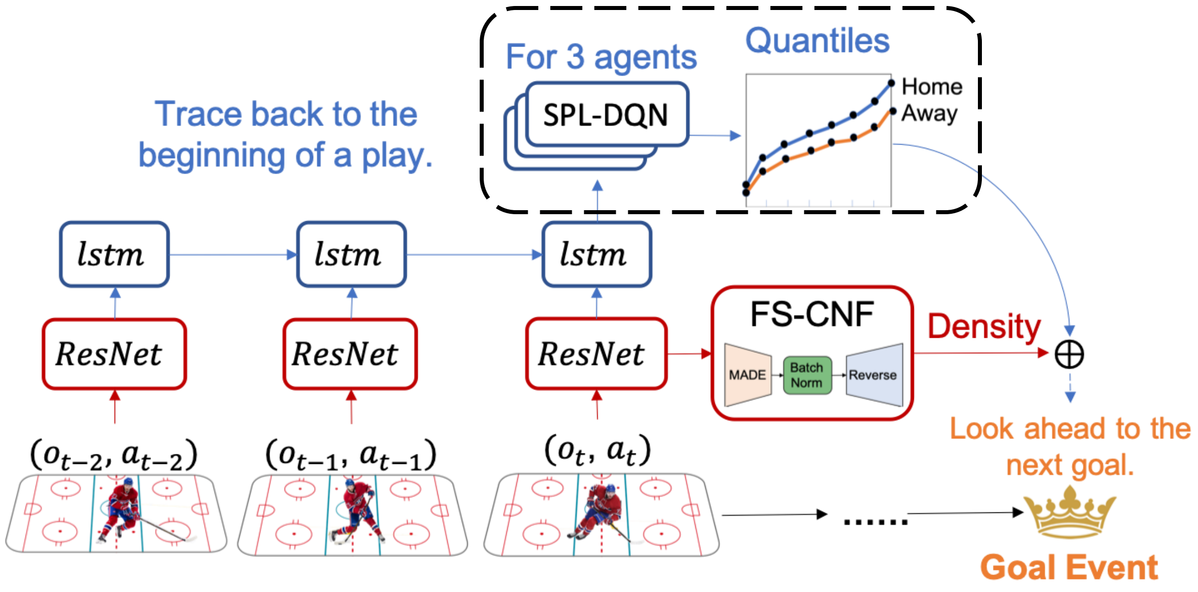
1) Learn the distribution of $Z_k(s_t, a_t)$, i.e., number of goals when a player performs action a_t in state s_t.

2) Represent $Z_k(s_t, a_t)$ by a uniform mixture of N supporting quantiles.

$$\hat{Z}_k(s_t, a_t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_{k,i}(s_t, a_t)}$$

($\theta_{k,i}$ estimates the ith quantile)

3) Distributional Bellman Operator
   Perform quantile regression to update



Treat Home team / Away team as two agents

Epistemic Uncertainty: Density Estimation

Value distribution in Distributional RL still contains epistemic uncertainty:

- In online learning: insufficient exploration
- In offline learning: insufficient data samples (our case)
- Common solution: density estimation, to distinguish in Distribution (InD) and out of distribution (OoD) datapoints
- May fail to capture epistemic uncertainty: feature collapse, i.e., map InD and OoD data to the same feature space

- Feature extractor should be **distance aware**: (intuition: if x is close to y, then f(x) close to f(y))
- Bi-Lipschitz condition

$$\beta_1 \|x_1 - x_2\|_I \geq \|f_\theta(x_1) - f_\theta(x_2)\|_F \geq \beta_2 \|x_1 - x_2\|_I$$

Upper bound ensures smoothness

Lower bound ensures sensitivity to distance

- Implement: residual network with spectral norm

Epistemic Uncertainty: Density Estimation

- Density Estimator for **Epistemic Uncertainty**

Feature Space Conditional Normalizing Flow (FS-CNF)

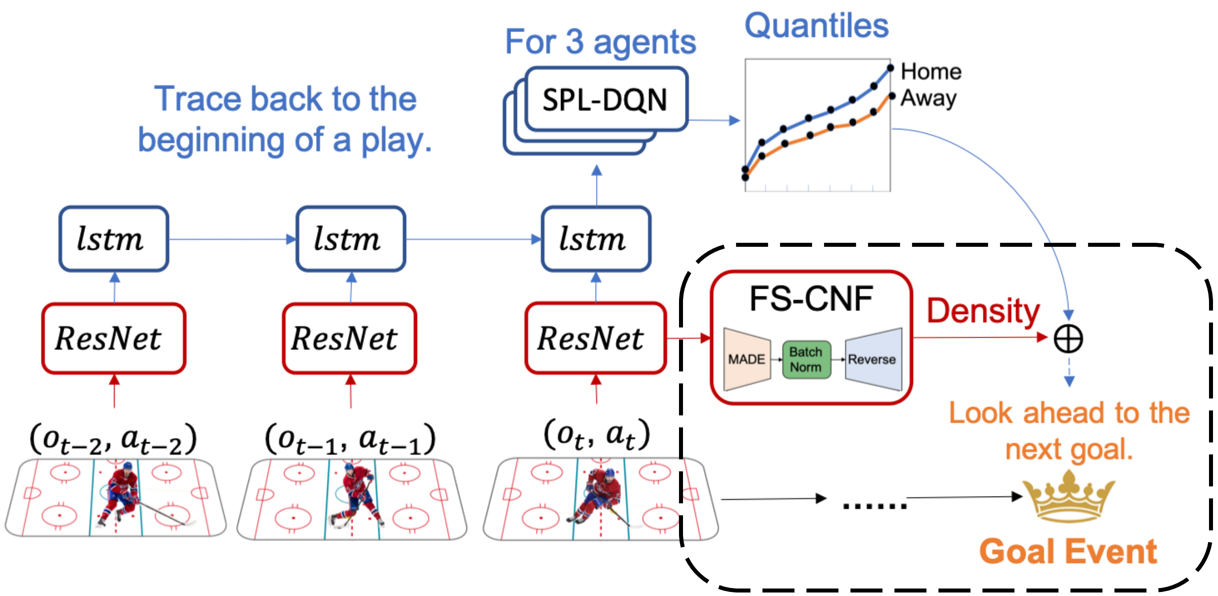
1) Feature Extractor.

To prevent feature collapse, the extractor is subjected to a bi-Lipschitz constraint:

$$\beta_1\|x_1 - x_2\|_I \geq \|f_\theta(x_1) - f_\theta(x_2)\|_F \geq \beta_2\|x_1 - x_2\|_I$$

Upper bound ensures smoothness

Lower bound ensures sensitivity to distance



2) Density Estimator.

Based on the extracted features, FS-CNF utilizes the Masked Auto-regressive Flow (MAF).

Player Evaluation

- **Risk-sensitive Impact Metric**

To understand how players respond to risk, we propose a Risk-sensitive Game Impact Metric (RiGIM)

Former

$$GIM_l = \sum_{(s,a)\in\mathcal{D}'} n(s,a,l) \times \phi(s,a)$$

where

$$\phi(s_{t+1}, a_{t+1}) = Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$$

number of times player L takes action a at state s

action impact

Ours

$$RiGIM_l(c) = \sum_{(s,a)\in\mathcal{D}'} n(s,a,l) \times \phi_k(s,a,c)$$

where

$$\phi_k(s_{t+1}, a_{t+1}, c) = [\hat{Z}_k^c(s_{t+1}, a_{t+1}) - \hat{Z}_k^c(s_t, a_t)]\mathbb{I}_{p(\cdot|z_E)\geq\epsilon}$$

confidence level c

(1-c) level quantile

density checker

Player Evaluation

- **Risk-sensitive Impact Metric**

To understand how players respond to risk, we propose a Risk-sensitive Game Impact Metric (RiGIM)

$$RiGIM_l(c) = \sum_{(s,a)\in\mathcal{D}'} n(s,a,l) \times \phi_k(s,a,c) \quad \text{where} \quad \phi_k(s_{t+1},a_{t+1},c) = \left[\hat{Z}_k^c(s_{t+1},a_{t+1}) - \hat{Z}_k^c(s_t,a_t)\right]\mathbb{I}_{p(\cdot|\boldsymbol{z}_E)\geq\epsilon}$$

(1-c) level quantile

- **Case Study: Player Ranking in Testing Games**

We rank players according to their RiGIM scores in the NHL testing games.

Table 1: Top 10 players with confidence 0.2.

Player Name	Position	Team	P	A	G	RiGIM
Jonathan Toews	C	CHI	10	5	5	14.72
Anze Kopitar	C	LAK	12	9	3	14.55
Vincent Trocheck	C	FLA	8	5	3	14.02
Tomas Hertl	C	SJS	12	8	4	13.97
John Tavares	C	TOR	12	3	9	13.92
Tyler Seguin	C	DAL	18	12	6	13.71
Leon Draisaitl	C	EDM	16	8	8	13.16
Aleksander Barkov	C	FLA	19	14	5	12.63
Sean Couturier	C	PHI	11	6	5	12.62
Nathan MacKinnon	C	COL	12	6	6	12.48

Risk seeking

Table 2: Top 10 players with confidence 0.8.

Player Name	Position	Team	P	A	G	RiGIM
Radek Faksa	C	DAL	6	3	3	2.74
Leon Draisaitl	C	EDM	16	8	8	2.51
John Klingberg	D	DAL	10	9	1	2.46
Esa Lindell	D	DAL	3	1	2	2.29
Connor McDavid	C	EDM	18	11	7	2.23
Tomas Hertl	C	SJS	12	8	4	1.93
Miro Heiskanen	D	DAL	5	3	2	1.86
Elias Pettersson	C	VAN	8	6	2	1.79
Tyler Seguin	C	DAL	18	12	6	1.78
Roope Hintz	LW	DAL	11	7	4	1.77

Risk averse

Defense man

Player Evaluation

Dataset

- Ice hockey from the National Hockey League, soccer from major European soccer leagues
- Over 9m events, over 4k games, over 6k players
- Event: (player who controls the puck or the ball)
  - player_id
  - action
  - other features

Type	Name	Range	
Ice Hockey	Spatial Features	X Coordinate of Puck	[-100, 100]
		Y Coordinate of Puck	[-42.5, 42.5]
		Velocity of Puck	$(-\infty, +\infty)$
		Angle between the puck and the goal	$[-3.14, 3.14]$
	Temporal Features	Game Time Left	[0, 3,600]
		Event Duration	$(0, +\infty)$
	In-Game Features	Score Differential	$(-\infty, +\infty)$
		Manpower Situation	{Even Strength, Shorted Handed, Power Play}
		Home or Away Team	{Home, Away}
		Action Outcome	{successful, failure}

Player Evaluation

Player Evaluation Performance: Correlations with Standard Measures (free online)

- Measure whether the metrics can form a comprehensive evaluation to a player's overall performance by computing the correlations between player ranking metrics and standard measures.

Table 4: Correlations with standard measures in the **ice hockey** games. The *success* measures are assist, goal, Game Winning Goal (GWG), Overtime Goal (OTG), Short-handed Goal (SHG), Power-play Goal (PPG), Point (P), Short-handed Point (SHP), Power-play Point (PPP), Time On Ice (TOI), and Shots (S). The *penalty* measure is Penalty Minute (PIM).

Methods	Assist	Goal	GWG	OTG	SHG	PPG	Point	SHP	PPP	TOI	S	PIM
+/−	0.181	0.189	0.187	0.028	0.071	0.063	0.206	0.119	-0.071	0.021	0.038	-0.014
EG	0.239	0.303	0.264	0.130	-0.053	0.163	0.322	0.023	0.226	0.153	0.534	-0.112
SI	0.237	**0.596**	**0.409**	0.123	0.095	0.351	0.452	0.066	0.274	0.224	0.405	0.138
VAEP	0.238	0.454	0.225	0.06	0.053	0.326	0.382	-0.0	0.321	0.086	0.362	0.027
T0-GIM	0.397	0.394	0.139	0.16	0.151	0.216	0.455	0.153	0.295	0.356	0.387	0.058
GIM	0.456	0.408	0.167	0.158	0.134	0.246	0.501	0.137	0.345	0.395	0.431	0.061
Na-RiGIM(0.5)	0.593	0.476	0.223	0.173	**0.152**	0.313	0.625	**0.175**	0.453	0.597	0.611	0.115
GDA-RiGIM(0.5)	0.591	0.475	0.221	0.174	**0.152**	0.315	0.623	0.174	0.452	0.593	0.609	0.113
RiGIM(0.5)	0.675	0.477	0.266	0.184	0.11	0.355	0.678	0.141	0.529	0.68	0.7	0.146
RiGIM(c^*)	**0.68**	0.477	0.269	**0.187**	0.107	**0.357**	**0.681**	0.141	**0.531**	**0.685**	**0.707**	0.147

We can choose some optimal confidence level c*=0.34 for ice hockey

Sensitivity to Risk: Correlations Conditioning on Different Confidence Levels

Measure whether RiGIM is sensitive to the risk by its correlations with the standard measures, where RiGIM is conditioned on a specific confidence level c (from 0 to 1)

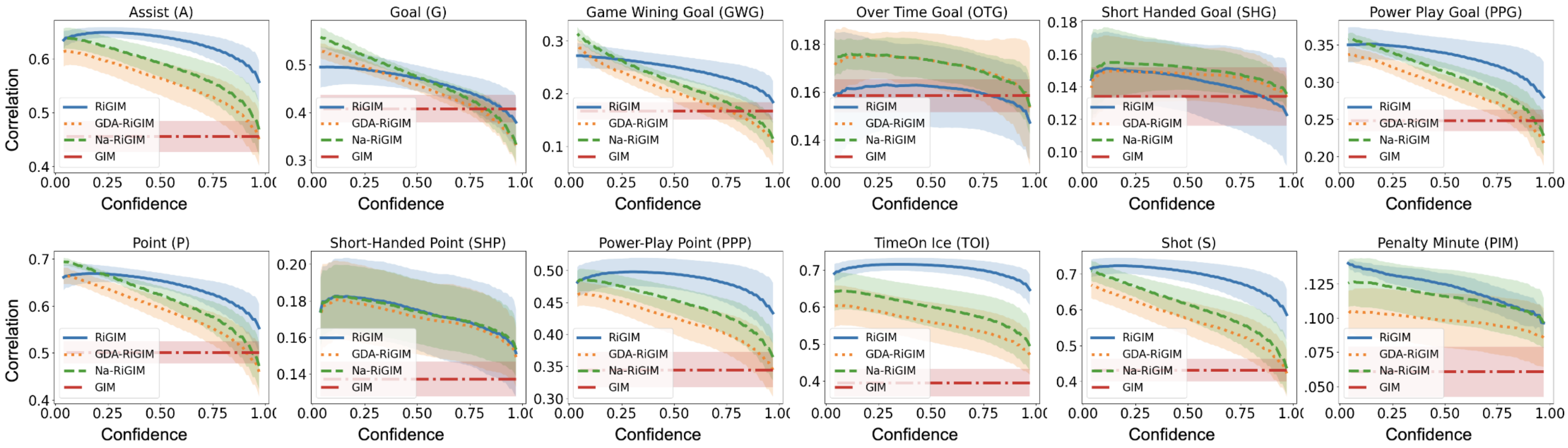


Figure 5: Correlations (Mean ± standard deviation) with success measures (the first 11 plots) and penalty measures (the last plot) at different confidence levels in **ice-hockey** games.

Question and Answering (Q&A)