11. Risk, Reward, and Reinforcement Learning in Ice Hockey Analytics

Sheng Xu, Oliver Schulte, Yudong Luo, Pascal Poupart, Guiliang Liu

Abstract

What makes many decisions in sports difficult is that they involve a trade-off between risk and reward. Actions such as taking a three-point shot, carrying a puck, or dribbling with a ball carry a higher risk of failure and require exceptional skill to pull off, but also bring a higher potential reward. This paper describes computational tools for *risk analytics* to model the risk inherent in the choices faced by teams and athletes. We leverage *distributional reinforcement learning (RL)* as a source of concepts and techniques for computational risk analytics. Distributional RL techniques allow us to model a dynamic distribution of outcomes for 1000+ games in the National Hockey League. We find strong evidence that strong teams take many risks (0.90 correlation between team season standing and team season standard/Gini deviation). For players, we also find strong evidence that stronger players take more risks (e.g., 0.86 correlation between a player's season goals and their value-at-risk metric).

11.1 Introduction: Taking Chances in Sports

Many decisions by athletes and coaches involve accepting higher risk for potentially higher rewards. A well-known example from basketball is whether to take a long-distance shot, which potentially nets three points, versus a shot from a shorter distance for two points. A more complex example from ice hockey is pulling the goalie—substituting an attacker for the goalie—when a team is trailing. Pulling the goalie earlier increases the chances of equalizing, but also increases the chances of the leading team scoring, which in practice decides the game immediately. While the decision-theoretically optimal choice is to *maximize the average success*, that is the average number of points and wins, several sports analysts have observed that players and coaches are often influenced by a secondary goal, which is to *minimize the probability of failure*, or generally the probability of bad outcomes. To illustrate the point in the basketball scenario, consider a player in a situation where the chance of scoring a three-pointer is 20% and the chance of scoring a two-pointer is 30%. Then the expected number of points is the same for each choice (namely 0.6). However, the probability of failure is 80% for the three-pointer and only 70% for the two-pointer; bad outcomes are less likely for the two-point shot. Now consider a different situation where the long-distance shot has a 25% chance of success. In this case, the expected number of long-distance points is 0.75 versus 0.6, and the optimal decision is to take the long-distance shot. However, the chance of failure is still 75% versus only 70% for the two-pointer, so a *risk-averse* player may still prefer the safer two-point shot.

Several sports analysts have argued that players and teams tend to take risk-averse decisions at the expense of their total success averaged over many games and match situations. Pelechrinis (2016) provides evidence that American football coaches aim to minimize the variance of expected points, rather than the expected points directly, perhaps to avoid public criticism for failure. Beaudoin and Swartz (2010) argued that trailing hockey teams should pull their goalies earlier to maximize winning chances. Indeed NHL teams have recently started pulling their goalies earlier.[1] Another piece of evidence for a trend towards more risk-taking is the rise of 3-point attempts in basketball, rising from 22.2% in the 2010-11 season to 39.2% in 2020-2021.[2] The trend towards riskier actions over time is evidence that risk-taking affords an advantage over more cautious tactics.

In this paper, we study risk-taking in the National Hockey League. We examine different ways to quantify how risky a decision is, including traditional notions such as the variance/standard deviation of outcomes, as well as the Gini deviation, an alternative variability concept well-suited to multi-modal distributions (Luo et al., 2023). From a technical viewpoint, *modeling risk requires modeling higher-order moments of the distribution of possible action outcomes.* To model the outcome distribution beyond its mean, we leverage recent work in *distributional reinforcement learning* (RL). Reinforcement learning is a branch of machine learning that studies how to act in sequential decision-making scenarios, such as we encounter in sports analytics. Reinforcement learning has developed an extensive set of methods for estimating expected outcomes from actions, known as prediction or policy evaluation methods (Sutton & Barto, 1998). Distributional reinforcement learning is a more recent development that provides methods for modelling the distribution of action outcomes. A recent paper by Liu et al. (2022) developed a distributional RL method for estimating the distribution of action outcomes for play-by-play (event) data. We utilize the computational tools from their work to estimate action outcome distributions from large play-by-play event data (1000+ games, 1M+ events).

We apply their framework to quantify and study the riskiness of actions by professional players in the NHL. The main question is how performance relates to risk-taking, for both teams and players. We examine three different variability concepts for quantifying the risk associated with a distribution of outcomes: standard deviation, Gini deviation, and value at risk. The risk impact of an action is the extent to which it increases/decreases the variability of the game outcomes for the acting player's team.

Our main findings are as follows: For *team performance*, the total risk of the actions taken by a team in a season displays a very high correlation with the team performance, measured by the number of total season points. Using standard deviation or Gini deviation as our risk measure, the correlation reaches 0.90. For the value-at-risk metric studied previously by Liu et al. (2022) the correlation is only 0.51, given a confidence level 0.2 that represents risk-seeking (see Section 11.6 for further details).

[1] https://www.nhl.com/news/coaches-room-paul-maclean-nhl-pulling-goalies-trend-301099580

[2] https://www.nba.com/news/3-point-era-nba-75

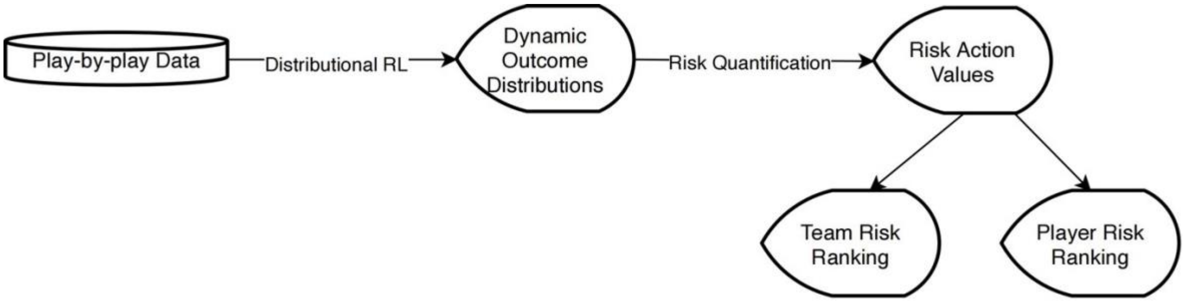

Figure 11.1: System Components in our Risk Analytics Framework.

For *player performance*, we use the total risk of the actions taken by a player in a season as a measure of the player's risk-taking. All player risk metrics show a high degree of temporal consistency, with their round-by-round totals essentially converging less than halfway through the season. The Gini deviation and standard deviation player metrics achieve substantial correlations of 0.56 and 0.51, respectively, with the player's total goals in a season. The value-at-risk metric achieves an even higher goal correlation of 0.86 (with the risk-seeking confidence level of 0.2). These results provide evidence that variability risk metrics are very good at predicting team success, but less suitable for predicting player success. However, because of the lack of a ground truth ranking for players, we do not consider this finding conclusive, and investigating risk-taking by players is a valuable direction for future research. While our study focuses on ice hockey data from the National Hockey League (NHL), our methods apply to any play-by-play dataset; see Liu et al. (2022) for an application to soccer data.

Paper Outline. Our paper is organized as follows. We begin with an overview of the rules of ice hockey and our play-by-play dataset. Then we review the background of reinforcement learning (RL), especially the distributional RL techniques for learning the distribution of action outcomes from play-by-play event data. Our discussion focuses on the main principles and intuitions; details on our learning methods may be found in the appendix and in the references. Given a dynamic distribution over future action outcomes at each point in a match, we define the risk impact of an action as the increase/decrease in the risk associated with the outcome distribution, after the action. The total risk impact of all actions is used to quantify risk-taking by teams and by players. Figure 11.1 summarizes our system components.

11.2 Hockey Rules and Hockey Data

NHL Rules. We give a brief overview of rules of play in the NHL (National Hockey League, 2014). NHL games consist of three periods, each 20 minutes in duration. A team has to score more goals than their opponent within three periods in order to win the game. Teams have five skaters and one goalie on the ice during even strength situations. Penalties result in a player sitting in the penalty box for 2, 4, 5 or 10 minutes and the penalized team will be shorthanded, creating a manpower differential between the two teams. The period where one team is penalized is called a powerplay for the opposing team with a manpower advantage. A shorthanded goal is a goal scored by the penalized team, and a powerplay goal is a goal scored by the team on the powerplay.

Dataset. In this paper, we use a play-by-play proprietary dataset constructed by Sportlogiq[3]. The data are constructed through a combination of computer vision and manual annotation. The dataset contains a total of 1196 games played between *October 3rd, 2018 and April 6th, 2019*. The training dataset for constructing our model contains 956 games (from October 3rd, 2018 to February 24th, 2019). Table 11.1 lists the features used in our analysis (Liu et al., 2022) and Figure 11.2 illustrates the adjusted coordinates in the Sportlogiq dataset. Figure 11.2 shows a schematic layout of the ice hockey rink. The units are feet. Adjusted Y- coordinates run from -42.5 at the bottom to 42.5. The goal line is at X = 89.

Table 11.1: The complete list of game features for the ice hockey dataset. The table utilizes adjusted spatial coordinates where negative numbers denote the defensive zone of the acting player and positive numbers denote the offensive zone.

	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/Action Duration	$(0, +\infty)$
	In-Game Features	Score Differential	$(-\infty, +\infty)$
		Manpower Situation	{Even Strength, Short-Handed, Power Play}
		Home or Away Team	{Home, Away}
		Action Outcome	{successful, failure}

The dataset records event data known as **play-by-play** data. Play-by-play data specifies the timing and location of actions, identifies the player responsible for each action, and includes the **contextual event features** outlined in Table 11.1. Table 11.2 lists the most frequent action types in

[3] https://sportlogiq.com

our dataset. To help visualize play-by-play data, Table 11.3 provides a partial sample.

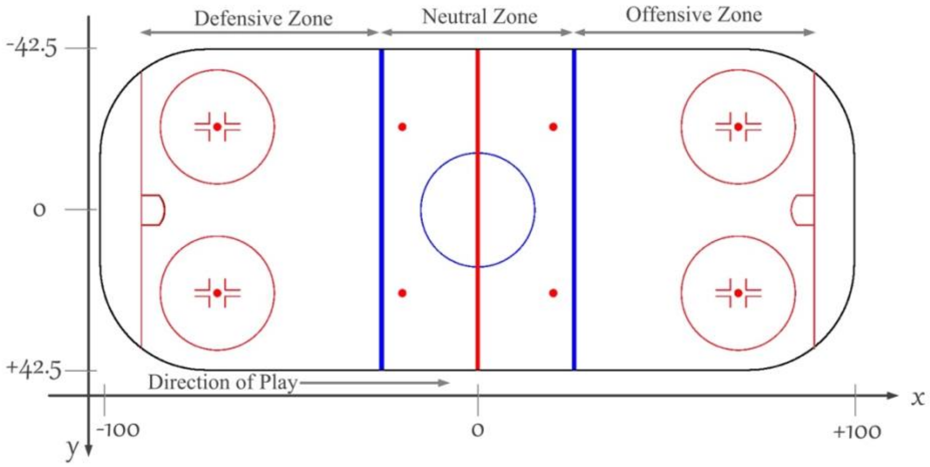

Figure 11.2: Rink layout with adjusted coordinates. Coordinates are adjusted so that for the team performing an action, its offensive zone is on the right.

Table 11.2: Definition of the most frequent Action Types in the Dataset.

Action	Description
Block	A block attempt on the puck's trajectory
Carry	Controlled carry over a blue line or the red center line
Check	When a player attempts to use his body to remove possession from an opponent
Dump in	When a player sends the puck into the offensive zone
Dump out	When a defending player dumps the puck up the boards without targeting a teammate for a pass
lpr	Loose puck recovery. The player recovered the puck as it was out of possession of any player
Offside	When a player is caught over the offensive blue line before their teammate brings the puck in
Pass	The player attempts a pass to a teammate
Puck protection	When a player uses their body to protect the puck along the boards
Reception	When a player receives a pass from a teammate
Shot	A player shoots on goal
Shot against	A shot was taken by the opposing team

Table 11.3: Sample Excerpt of Play-By-Play Data.

gameId	playerId	period	teamId	xCoord	yCoord	Manpower	Action Type
849	402	1	15	-9.5	1.5	even	lpr
849	402	1	15	-24.5	-17	even	carry
849	417	1	16	-75.5	-21.5	even	check
849	402	1	15	-79	-19.5	even	puckprot
849	413	1	16	-92	-32.5	even	lpr
849	413	1	16	-92	-32.5	even	pass
849	389	1	15	-70	42	even	block
849	389	1	15	-70	42	even	lpr
849	389	1	15	-70	42	even	pass
849	425	1	16	-91	34	even	block
849	395	1	15	-97	23.5	even	reception

11.3 Markov Game Models for Sports

Reinforcement learning provides a rich toolkit for estimating the chances of future success for strategic agents. Schulte (2022) gives a short accessible introduction to applying RL in sports analytics. For single-agent problems, RL is based on the fundamental Markov decision process model. Generalizing Markov decision process to multiple decision makers leads to the Markov game model. Markov game models have been developed for several sports, such as ice hockey, soccer, and American football (Chan et al., 2020; Liu et al., 2020; Liu & Schulte, 2018). We utilize the ice hockey model of Liu and Schulte (2018).

11.3.1 Markov Game Model for NHL Ice Hockey

Similar to (Liu & Schulte, 2018), we apply the Markov Game Framework to model the play dynamics for sports games. A Markov Game (Littman, 1994), sometimes called a stochastic game, is defined by a set of states $\mathcal{S}$, and a collection of action sets $\mathcal{A}$, one for each agent in the environment. State transitions are controlled by the current state and a list of actions, one action from each agent. For each agent, there is an associated reward function mapping a state transition to a reward. An overview of how a hockey Markov Game model fills in this schema is as follows.

- There are two agents, *Home* and *Away*, representing their respective teams.
- The **action** a_t denotes the movements of players who control the puck. Our model applies a discrete action vector using a one-hot representation.
- An **observation** is a feature vector x_t specifying a value of the features listed in Table 11.1 at a discrete time step t . We use the complete sequence $s_t \equiv (x_t, a_t, x_{t-1}, \ldots, x_0)$ to represent the **state** (Mnih et al., 2015).

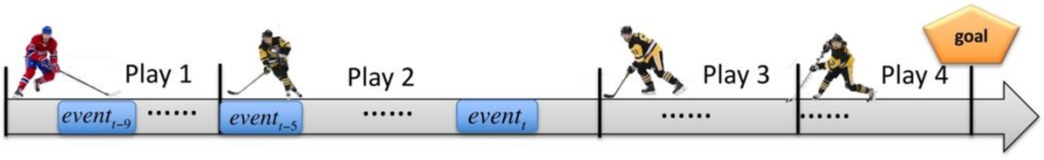

Figure11. 3: A hockey match is segmented into goal-scoring episodes.

- Since we have two agents, we have two **reward functions**, one for the home team and one for the away team. Conceptually, the reward at time t is 1 for a team that scores a goal at time t, 0 if there is no goal; we write $\text{goal}_{t,Home}$, $\text{goal}_{t,Away}$. For example, $\text{goal}_{t,Home} = 1$ indicates that the home team scores at time t. It is technically useful to introduce a virtual "none" agent for the eventuality that neither team scores until the end of a game. If neither team scores at the end of the match, we write $goal_{T,\,Neither} = 1$ where T is the last time step.

The **expected goal model** $R_k(s_t, a_t) = P\big(\text{goal}_{t,k} = 1 | s_t, a_t\big)$ specifies the probability that a team scores a goal after an action in a given match state. This model makes the **Markov assumption**, which implies that the state information available at time t is informative enough that scoring chances can be estimated based on the current state only, independent of the current time t and previous states. Technically, the Markov game model is stationary and the goal scoring probability is independent of the current time index t. Similarly, transition probabilities $P(s_{t+1}, a_{t+1} | s_t, a_t)$ are assumed to be stationary and depend on the current match state only. This means that for a given match state s_t and action a_t at time t, the *dynamics of NHL play define a distribution over future game trajectories that depends only on the current state and action.*

11.3.2 The Expected Value Function

We divide a sports game into **goal-scoring episodes**, so that each episode: 1) starts at the beginning of the game, or immediately after a goal, and 2) terminates with a goal or at the end of the game (s_H). Episodes extend through period breaks.

A key quantity in reinforcement learning is the *expected reward* with respect to future trajectories. Given our binary reward (score goal or not), the expected reward for a team k is the chance of scoring the next goal, denoted as $Q_k(s_t, a_t)$. To explain the basic RL approach to learning a Q-function, consider first the expected reward for a bounded *horizon*, that is, a fixed look-ahead length H . The chance of scoring within the next *H* steps is then defined by the expression:

$$Q_k^H(s_t, a_t) = \sum_{h=0}^{H} \gamma^h \, P\big(\text{goal}_{t+h,k} = 1 | s_t, a_t\big) \quad \mathbf{(1)}$$

Following previous studies (Liu & Schulte, 2018; Liu et al., 2020), we set $\gamma = 1$. In this case the Q_k^H value simply denotes the probability that team k scores a goal within H steps.

The Q-value satisfies an important recurrence relation known as the dynamic programming update:

$$Q_k^{H+1}(s_t, a_t) = R_k(s_t, a_t) + \gamma E_{s_{t+1}, a_{t+1} \sim P(s\prime, a\prime | s_0 = s_t, a_0 = a_t)}\big[Q_k^H(s_{t+1}, a_{t+1})\big] \quad (2)$$

The principle behind dynamic programming is that a team scores a goal in $H + 1$ steps if and only if they (1) score immediately or (2) take another step and then score within H steps. The value iteration algorithm uses dynamic programming to estimate Q-values from a dataset $\mathcal{D}$ of observed trajectories as follows.

1. Initialize the Q-values for $H = 0$ with the expected goal model R .
2. Iteratively apply Equation (2) through $H = 1, H = 2, \ldots,$ until convergence where $Q_k^{H+1}(s_t, a_t) = Q_k^H(s_t, a_t)$ for each team and state-action pair. We denote the **convergent Q-value** as $Q_k(s_t, a_t)$.

For continuous state spaces, such as we have in ice hockey, the expectation $E_{s_{t+1}, a_{t+1} \sim P(s', a' | s_0 = s_t, a_0 = a_t)}$ can be estimated by averaging over all transitions $(s_t, a_t; s_{t+1}, a_{t+1})$ observed in the data set. In the NHL model of Schulte et al. (2017b), convergence occurred with a lookahead of $H = 13$. The fact that expected values in RL incorporate lookahead means that they can capture the medium-term effects of actions on goals scoring.
With an unbounded lookahead $H \to \infty$, under mild conditions value iteration converges to a fixed point that satisfies:

$$Q_k(s_t, a_t) = R_k(s_t, a_t) + \gamma E_{s_{t+1}, a_{t+1} \sim P(s_{t+1}, a_{t+1} | s_t, a_t)}[Q_k(s_{t+1}, a_{t+1})] \qquad (3)$$

which is known as the **Bellman equation** for policy evaluation. In the Appendix 11.10.1 we discuss how the Bellman equation can be applied to learn a neural net model of the Q-function.

Remark. For readers familiar with reinforcement learning models, we briefly situate our NHL model with respect to other RL models. Other readers can skip this paragraph without loss of continuity.Our learning setting is *off-line* learning where we learn a value function from a dataset without executing actions; that is, our problem is prediction not control. In the off-line perspective, the observed actions can be treated as another feature similar to states. Formally, what we have defined is a *Markov reward process* in an expanded state space $\mathcal{S} \times \mathcal{A}$ where an expanded state is a pair (s, a) (cf. (Sutton& Barto, 1998, Ch.6.4)). The Q-function as we have defined it is the value function of this Markov reward process. We have used the Q-notation, rather than V for value function, because its meaning is the same as in policy evaluation: the expected cumulative reward for an agent given a current action and a current state. An equivalent model would be to first estimate a policy π_{Home} for the home team and another policy π_{Away} for the away team. For example, if in state s the home team passes the puck 30% of the time, we might estimate $\pi_{Home}(pass|s) = 30\%$. The Q-function as we have defined it represents $Q^{\pi_{Home}, \pi_{Away}}$, which is the Q-function of the NHL Markov game where the home and the away team follow the behavioral policies with action frequencies shown in the data. While the Markov reward model for off-line data is perhaps less familiar than the policy evaluation formulation, we use it because it is conceptually simpler and in fact fits naturally the position of the sports analyst who is passively watching the matches: decisions by the players are events for the sports analyst to analyze, not choices to control. For more discussion of reducing off-line Markov game analysis to other RL models please see (Luo et al., 2020).

11.3.3 Learning Reward Distributions

Distributional RL learns the distribution of the random variable $Z^k(s_t, a_t)$ that returns the sum of (discounted) rewards for future episode trajectories starting with state s_t and a_t (Bellemare et al., 2017). Therefore the Q-value is the expectation of the Z variable: $Q_k(s_t, a_t) = E(Z_k(s_t, a_t))$. Similar to the

Q-value, the distribution of the total rewards Z follows the **distributional Bellman equation**:

$$Z_k(s_t, a_t) :\triangleq R_k(s_t, a_t) + \gamma E_{s_{t+1}, a_{t+1} \sim P(s_{t+1}, a_{t+1} | s_t, a_t)}[Z_k(s_{t+1}, a_{t+1})] \quad (4)$$

where $\mathrm{X} :\overset{\Delta}{=} \mathrm{Y}$ indicates that random variables X and Y follow the same distribution. Given a computationally tractable representation of the Z_k distributions, we can iteratively apply the Bellman update Equation (4) to update the distribution for different look-ahead lengths $H = 1, H = 2, \ldots$, until we arrive at a **convergent distribution** Z_k.

Computational Representation. To make the distributional Bellman equation operational, the question is how to choose a computationally tractable representation that supports learning. One option is to choose a parametric family, such as a Gaussian distribution. The issue with such parametric families is that they are typically unimodal. Unimodal distributions are not appropriate for the complex dynamics of sports, where events typically have a high branching factor, with different branches corresponding to different modes. For example, if a player attempts a pass, three different possible outcomes are that the pass is 1) intercepted, 2) reaches the intended recipient, or 3) turns into a loose puck. Each of these outcomes has alternative subsequent events, etc. Different possible event sequences determine different scoring probabilities, leading to a highly multi-modal distribution.

Bellemare et al. (2017) proposed modelling reward distributions using **quantile regression**; Liu et al. (2022) applied quantile regression to ice hockey and soccer data.

The quantile-regression (QR)-DQN method represents the conditional distribution of Z by a uniform mixture of N supporting quantiles as $\widehat{Z_k}(s_t, a_t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_{k,i}(s_t, a_t)}$, where $\theta_{k,i}$ estimates the quantile at the quantile level (or quantile index) $\tau_i = i/N$ for $1 \leq i \leq N$ and $\delta_{\theta_{k,i}}$ denotes a Dirac distribution at $\theta_{k,i}$. For example, suppose we take $N = 4$ and estimate quantiles at 25%, 50%, 75%, 100% as 0.1, 0.4, 0.7, 0.9. Then the cumulative density function (cdf) of $\widehat{Z_k}(s_t, a_t)$ has 25% of values at 0.1 or less, 50% (the median) of values at 0.4 or less, 75% of values at 0.7 or less, and 100% of values at 0.9 or less (so none above 0.9). Within each pair of quantiles, the cdf is approximated as uniform over the quantiles (e.g., the cdf has value 0.1 between 0 and 25%).[4]

Given a fixed number N of target quantiles ($N = 4$ in our example), we can train a neural network to take as input a state-action pair (s, a), and output 4 numbers corresponding to the quantiles. By increasing the number N, this procedure provides a non-parametric approximation to the cumulative density function of $\hat{Z}$ and therefore to the distribution of $\hat{Z}$. For further details please see (Liu et al., 2022).

Choice of Outcome Variable: Expected Goals vs. Actual Goals. We obtained good empirical results with actual goals as rewards. As a further refinement, we follow (Decroos et al. 2017) and decompose goal scoring probabilities into the probability of managing a shot and the probability that a goal leads to a shot:

$$P(goal_{t,k} | s_{t_0}) = P(goal_{t,k} | s_t, shot) \times P(\mathrm{s_t}, \mathrm{shot} | \mathrm{s_{t_0}})$$

which can be read as saying that probability of scoring a goal from an initial state s_{t_0} is the probability of managing a shot times the probability of the shot leading to a goal. This equation is true in hockey because the only way to score a goal is to first take a shot. In our application of distributional RL, we take the goal scoring probabilities $P(goal_{t,k} | s_t, shot)$ as the outcomes (virtual rewards) whose

[4] We have slightly simplified notation compared to Bellemare et al. (2017) where a quantile is associated with the midpoint of a bins, rather than the higher endpoint if the bin.

distribution is to be modelled.

The motivation for using the shot-goal decomposition is as follows:

1. A team can largely control whether they achieve a shot, whereas the success of the shot depends on factors such as the skill of the opposing goalie that are less under the control. So a model of team/player strength should reward teams for managing shots.
2. Shots are sparse but not as sparse as goal.
3. With actual goals as rewards, our outcome variable Z_k is binary and the outcome distribution is basically a Bernoulli distribution. With expected goals as rewards, the outcome variable Z_k ranges over the interval [0,1], and the outcome distribution is an informative distribution over goal scoring probabilities.

So far, we have described how to build a machine learning model that estimates the distribution of possession outcomes for a given team, given a specific time and context in a match. We now show how to apply the model to gain analytical insights for a sport. Specifically, we discuss how to evaluate risk-taking by teams, the riskiness of actions, and ranking players by how risky their actions are.

11.4 Risk Measures for an Outcome Distribution

Luo et al. (2023) consider in depth the properties of different measures of risk for a distribution of outcomes. In this study we employ three of these measures: standard deviation, Gini deviation, and Value-at-Risk (VaR). Variance, the square of standard deviation, and VaR are commonly used risk measures in portfolio analysis. Gini deviation is recommended by Luo et al. (2023) for multi-modal distributions. The formal definitions are as follows.

Definition 1 (Risk Measures)**.** *For a random variable Z, let Z_1 and Z_2 be two i.i.d. copies of Z, that is, Z_1 and Z_2 are independent and follow the same distribution as Z.*

- *The variance is defined as* $\mathbb{V}[Z] = \frac{1}{2}\mathbb{E}[(Z_1 - Z_2)^2]$
- *The standard deviation is the square root of the variance* $\mathrm{STD}[Z] = \sqrt{\mathbb{V}[Z]}$
- *The Gini deviation is defined as* $\mathbb{D}[Z] = \frac{1}{2}\mathbb{E}[|Z_1 - Z_2|]$.

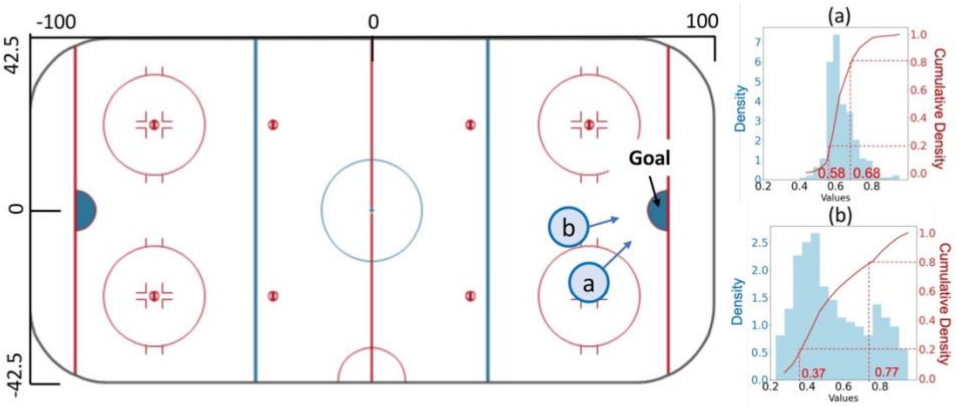

Figure 11.4: The predicted distribution of future goals in an ice hockey game between Blues and Coyotes, 2018-19 NHL season. The shots are made in the positions (a) and (b). Next-goal scoring distributions (a) and (b) have *the same expectation* (around 0.6), but the first shot has a much lower variance and Gini deviation of outcomes. Shot (a) also displays a larger risk-averse estimate (at the confidence 0.8, we find a larger next-goal chance with 0.58 > 0.37) and a smaller risk-seeking estimate (at the confidence 0.2, we find a smaller next-goal chance of 0.68 < 0.77).

Thus the Gini deviation replaces the L2 norm of variance by the L1 norm. The definition **value-at-risk** (VaR) depends on the choice of a **confidence level** $c \in (0,1]$. The VaR for level c is defined as the $(1-c)^{th}$ quantile in the distribution. Thus in the hypothetical example from above, the value at risk for *c* = 25% is 0.7, and for *c* = 50% it is 0.4. Intuitively, VaR provides a kind of worst-case analysis with respect to a user-controlled risk-level. For example, choosing *c* = 0.8 corresponds to *risk-aversion* since it focuses on bad outcomes. In contrast, choosing *c* = 0.2 corresponds to *risk-seeking* with better sensitivity to positive outcomes. Figure 11.4 illustrates how different risk concepts apply to different outcome distributions.

Computing Risk Measures from Quantile Regression VaR is defined in terms of quantiles and thus can be naturally computed from a quantile regression model. Given a quantile representation, we can estimate the variance and Gini deviation as follows (Luo et al., 2023).

$$\mathbb{V}\big[\hat{Z}_k\big] \approx \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\big(\theta_{k,i}-\theta_{k,j}\big)^2 \qquad (5)$$

$$\mathbb{D}\big[\hat{Z}_k\big] \approx \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\big|\theta_{k,i}-\theta_{k,j}\big| \qquad (6)$$

Figure 11.5 illustrates the Q-values and standard deviations that the trained model assigns during a game between the Flyers and Maple Leafs.

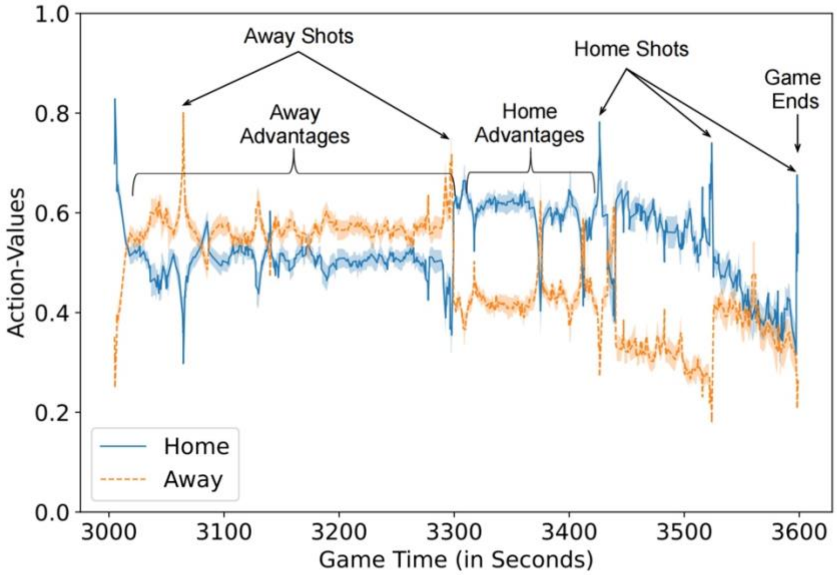

Figure 11.5: Illustrating the dynamic distribution of next-goal scoring chances by showing the corresponding mean ± standard deviation of the action values at each time step in a match between the Flyers (Home team) and the Maple Leafs (Away team) on March 15, 2019.

11.5 Measuring the Riskiness of an Action

Using the techniques described in the previous section, we can compute from an estimated outcome distribution $\hat{Z}_k(s_t, a_t)$ a risk measure $\rho_k(s_t, a_t)$, where ρ is one of the risk measures described above (standard/Gini deviation, VaR(c)). A simple approach would be to measure the riskiness of an action in a match state simply by the risk measure $\rho_k(s_t, a_t)$. The problem with this approach is that it measures the general match context of the action, rather than the specific *impact* of the action. For example if a player makes a pass when his team has an empty net, the risk measure will be high regardless of how risky his pass is. Intuitively, the pass is taking place in a risky place at a risky time, but may itself not contribute to a team's risk.

The same issue arises with respect to expected action values (i.e., Q-values): A team playing against an empty net has a high chance of scoring the next goal, but a particular action by a player may not be increasing his team's scoring chances beyond playing an empty net. Routley and Schulte (2015) proposed to address this issue by computing the **action impact**, which is measured by how much an action *changes* the scoring chances of the team in possession. The action **goal impact** is defined as follows for an action a_{t+1} and state s_{t+1} comprising the $t+1$ -th event:

$$impact_k(s_{t+1}, a_{t+1}) = [Q_k(s_{t+1}, a_{t+1}) - Q_k(s_t, a_t)]\mathbb{1}_{p(s_{t+1}, a_{t+1}) \geq \epsilon} \qquad (7)$$

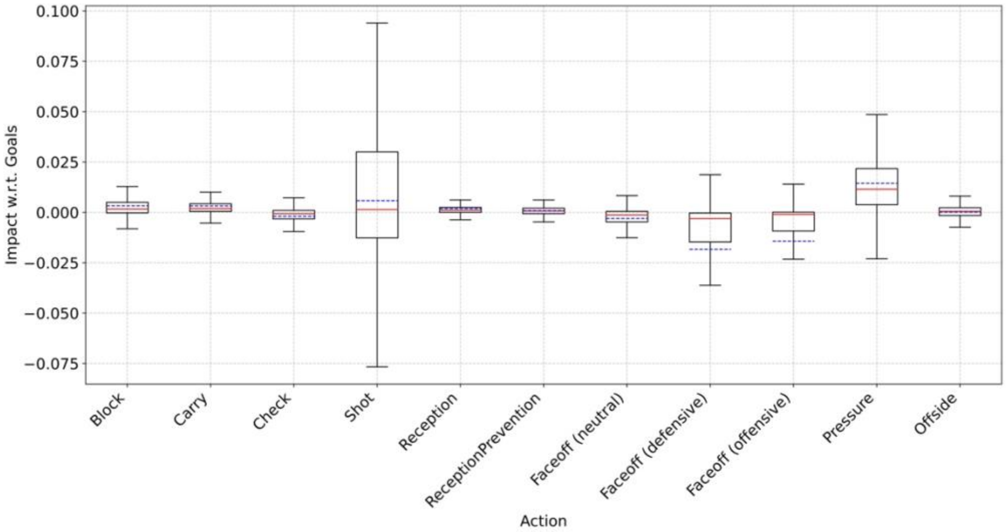

Figure 11.6: Box Plot for the risk impact of an action on the standard deviation of next-goal scoring chances. The red line indicates the median value and the blue line indicates the mean value. Each point represents the impact of the given action in a state (outliers removed).

Here p is a density estimator for state-action pairs, and $\mathbb{1}_{p(\cdot) \geq \epsilon}$ is an indicator value that returns 0 if the probability of the event falls below a threshold ϵ. The idea is to filter out rare events because the Q-value estimates for rare events are often biased due to small sample sizes. Using a discrete Markov game model, Routley and Schulte (2015) removed all state-action pairs from consideration that occur less than 10 times in the data. For our continuous state space, we follow Liu et al. (2022) and eliminate anomalous events with $p < 20\%$. Anomaly filtering is not an essential component of our risk analysis framework.

We can adapt the goal impact approach by measuring how much an action changes the risk of an action, which we call the **risk impact**:

$$Rimpact_k(s_{t+1}, a_{t+1}) = [\rho_k(s_{t+1}, a_{t+1}) - \rho_k(s_t, a_t)]\mathbb{1}_{p(s_{t+1},a_{t+1})\geq\epsilon} \qquad (8)$$

where ρ is one of our risk measures (standard/Gini deviation, value at risk).
Figures 11.6 and 11.7 show the box plots for the standard/Gini deviation impacts for different actions. Game-changing events such as shots tend to have a high impact on the variability of team outcomes. Defensive faceoffs tend to decrease a team's risk, likely because they follow a successful defense. It is interesting to note that exerting pressure tends to increase a team's risk. Box plots for value-at-risk are in the appendix.

11.6 Team Performance and Team Risk-tasking

We apply our outcome distribution model to teams, by evaluating how much risk a team takes on aggregate. We start with teams because we can use the total team performance over a season as a ground-truth metric of team success. *The main question in this section is whether risk-taking by a team correlates with team success.* To quantify risk-taking by teams, we add up the risk impact of the team's total actions.

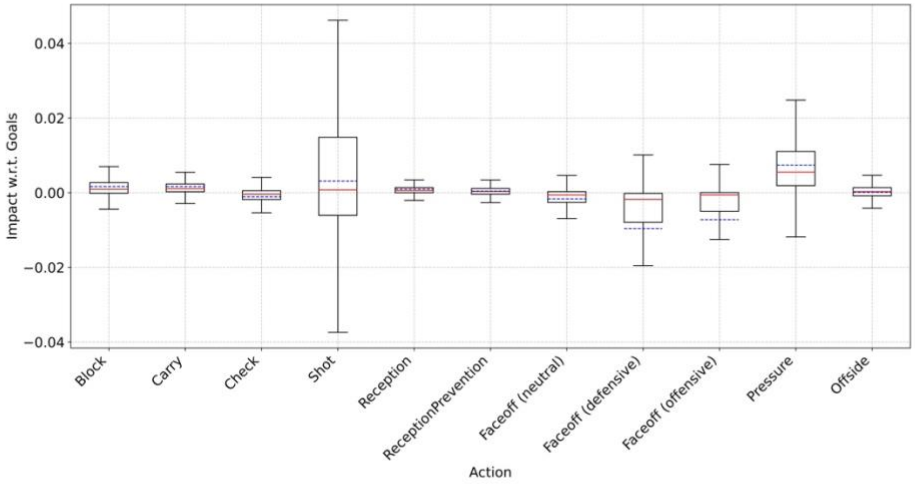

Figure 11.7: Box Plot for the risk impact of an action on the Gini deviation of next-goal scoring chances. The red line indicates the median value and the blue line indicates the mean value. Each point represents the impact of the given action in a state (outliers removed).

For a dataset $\mathcal{D}$, let g, t be a generic instance for event number t in game number g . Also let $team_{gt}$ be the team in possession at event t in game g, and similarly for the state s_{gt} and the action a_{gt}. Then the total team risk impact for team T in the dataset is given by

$$RIM_k = \sum_{g,t:team_{gt}=T} Rimpact_{k_t}(s_{gt}, a_{gt}) \qquad (9)$$

where k_t denotes the appropriate agent at time t (Home or Away). Equation (9) shows how to define a team risk impact metric for each risk measure, which we abbreviate as follows: StdRIM = risk impact for standard deviation, GdRIM = risk impact for Gini deviation, RIM(0.2) = risk impact for VaR with risk-seeking confidence level 0.2, RIM(0.8) = risk impact for VaR with risk-averse confidence level 0.8. RIM(c) is denoted as RiGIM(c) by Liu et al. (2022) using the same notation.

A team's season league standing is determined by the number of points the team earns in each match. We therefore measure the correlation between a team's season risk impact and season total points as a measure of team performance. As Figures 11.8 and 11.9 show, both the standard and Gini deviations provide measures of risk-taking that are excellent predictors of team performance. This shows that *stronger teams take more risks.* Table 11.4 provides the Pearson correlations between team total points and team risk metrics. Value-at-risk metrics are substantially less informative about team performance, especially at the high confidence level of 0.8. Table 11.5 shows the top 10 teams in the league and their risk metrics.

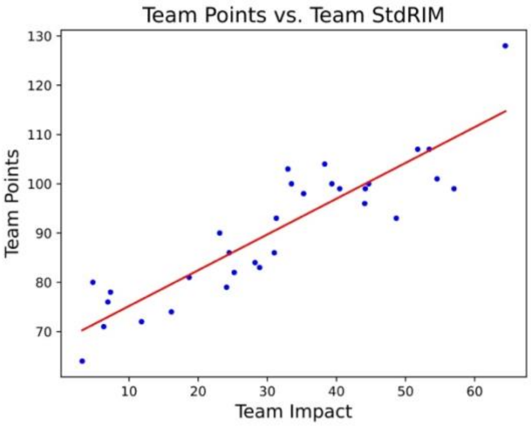

Figure 11.8: Team Points Vs. Team StdRIM.

Table 11.4: Correlations between a team's risk-impact metric and their season totals. The standard/Gini deviations show a very high predictive ability for team season performance, which shows that stronger teams take more risks.

StdRIM	GdRIM	RIM(0.2)	RIM(0.8)
0.90	0.90	0.51	-0.24

11.7 Ranking Hockey Players by Risk-taking

As with teams, we use the total risk impact of a player's actions to evaluate their risk-taking. Let pl_{gt} be the player in possession at event t in game g . Then the **total player risk impact** for player l in the dataset is given by

$$RIM_l = \sum_{g,t:pl_{gt}=l} Rimpact_{k_t}(s_{gt}, a_{gt}) \qquad (10)$$

where we use the same notation as in Section 6. Tables 11.6 and 11.7 show the top 20 players according to their risk ranking. Applying the "eye test", risk-seeking VaR(0.2) identifies many stars, such as Connor McDavid, Leon Draisaitl, Sidney Crosby. The standard deviation metric also identifies several stars, such as Alex Ovechkin and Johnny Gaudreau. But it also highlights several less-heralded players, such as Jason Zucker and Jaden Schwartz. Another difference is that the standard deviation metric shows a bias towards forwards, whereas the Var(0.2) risk impact metric includes some centres. The top 20 tables for the other risk metrics are in the appendix.

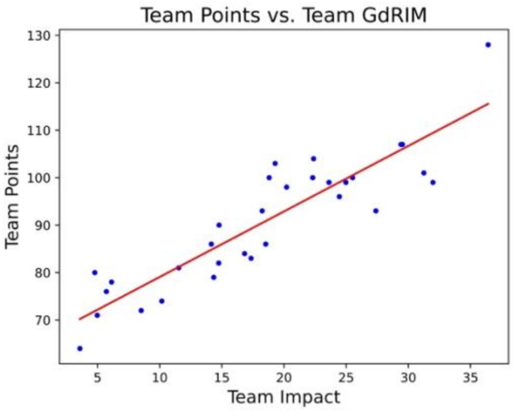

Figure 11.9: Team Points Vs. Team GdRIM.

Table 11.5: Top 10 teams with StdRIM and GdRIM based on the entire season. The real rank is based on Total Points. The predicted rank is based on the risk impact metrics (both agree on the ranking), which measures a team's total risk taking over the season.

Team Name	Predicted Rank	Real Rank	Total Points	StdRIM	GdRIM
Lightning	1	1	128	64.43	36.42
Blues	2	12	99	57.01	31.98
Sharks	3	6	101	54.56	31.25
Flames	4	2	107	53.43	29.53
Bruins	5	3	107	51.76	29.41
Golden Knights	6	16	93	48.66	27.39
Maple Leafs	7	7	100	44.67	25.52
Hurricanes	8	11	99	44.17	24.98
Canadiens	9	14	96	44.09	24.45
Jets	10	10	99	40.47	23.62

Table 11.6: Top 20 players according to the risk metric value-at-risk with a risk-seeking confidence level 0.2 (i.e., RIM(0.2)) based on the entire season.

Player Name	Position	Team	P	A	G	RIM(0.2)
Nikita Kucherov	RW	TBL	128	87	41	61.12
Mitchell Marner	RW	TOR	94	68	26	60.81
Johnny Gaudreau	LW	CGY	99	63	36	59.71
Patrick Kane	RW	CHI	110	66	44	56.55
Brad Marchand	LW	BOS	100	64	36	53.34
Mark Stone	RW	VGK	73	40	33	51.29
Connor McDavid	C	EDM	116	75	41	51.00
Leon Draisaitl	C	EDM	105	55	50	50.16
Timo Meier	RW	SJS	66	36	30	49.67
Blake Wheeler	RW	WPG	91	71	20	48.96
Sidney Crosby	C	PIT	100	65	35	48.56
Jonathan Huberdeau	LW	FLA	92	62	30	48.19

Kyle Connor	LW	WPG	66	32	34	47.85
Artemi Panarin	LW	CBJ	87	59	28	47.34
Evgenii Dadonov	RW	FLA	70	42	28	45.83
Cam Atkinson	RW	CBJ	69	28	41	45.70
Matthew Tkachuk	LW	CGY	77	43	34	45.32
Brendan Gallagher	RW	MTL	52	19	33	44.95
Jake Guentzel	LW	PIT	76	36	40	44.58
Brandon Saad	LW	CHI	47	24	23	43.69

Table 11.7: Top 20 players with StdRIM based on the entire season.

Player Name	**Position**	**Team**	**P**	**A**	**G**	**StdRIM**
Johnny Gaudreau	LW	CGY	99	63	36	13.07
Patrick Kane	RW	CHI	110	66	44	12.04
Nikita Kucherov	RW	TBL	128	87	41	11.68
Alex Ovechkin	LW	WSH	89	38	51	11.30
Mitchell Marner	RW	TOR	94	68	26	11.14
Cam Atkinson	RW	CBJ	69	28	41	11.05
Timo Meier	RW	SJS	66	36	30	10.77
Vladimir Tarasenko	RW	STL	68	35	33	9.31
Matthew Tkachuk	LW	CGY	77	43	34	9.30
Brad Marchand	LW	BOS	100	64	36	9.12
Jaden Schwartz	LW	STL	36	25	11	9.10
Brendan Gallagher	RW	MTL	52	19	33	9.04
David Pastrnak	RW	BOS	81	43	38	9.01
Kyle Connor	LW	WPG	66	32	34	8.95
Filip Forsberg	LW	NSH	50	22	28	8.72
Josh Anderson	RW	CBJ	47	20	27	8.30
Mikko Rantanen	RW	COL	87	56	31	8.28
Evgenii Dadonov	RW	FLA	70	42	28	7.91
Jason Zucker	LW	MIN	42	21	21	7.85
Jonathan Huberdeau	LW	FLA	92	62	30	7.73

A difficulty in evaluating player rankings is that unlike with teams, there are few suitable ground-truth metrics for performance (Franks et al., 2016). We follow previous work (Liu & Schulte, 2018; Decroos et al., 2019) and consider the correlations between our risk metrics and other meaningful player statistics such as goals, assists, and points (= goals + assists). Figure 11.10 plots round-by-round correlations between these metrics and the deviation risk impact metrics. For each round in the season, for each player, we compute their total risk impact so far (e.g., standard deviation impact over all games up to round 30), and correlate it with their statistics (e.g., total goals scored up to round 30). We observe a substantive correlation for goal-based statistics, reaching 0.51 for StdRIM and 0.56 for GdRIM at season's end. Note that the correlation is already relatively high after about 30 rounds, less than half-way through the season. This means that *the risk-impact metrics have high predictive power for future player performance.* The auto-correlation plot in the bottom right directly confirms the temporal consistency of the risk metrics. This plot correlates the value of the risk metric after *n* rounds with the final season value. We see that already after 25 rounds, a player's risk impact observed so far predicts their final risk impact with correlation above 0.8. The strong temporal auto-correlation is evidence that risk-taking measures capture a stable player's characteristics (Pettigrew, 2015).

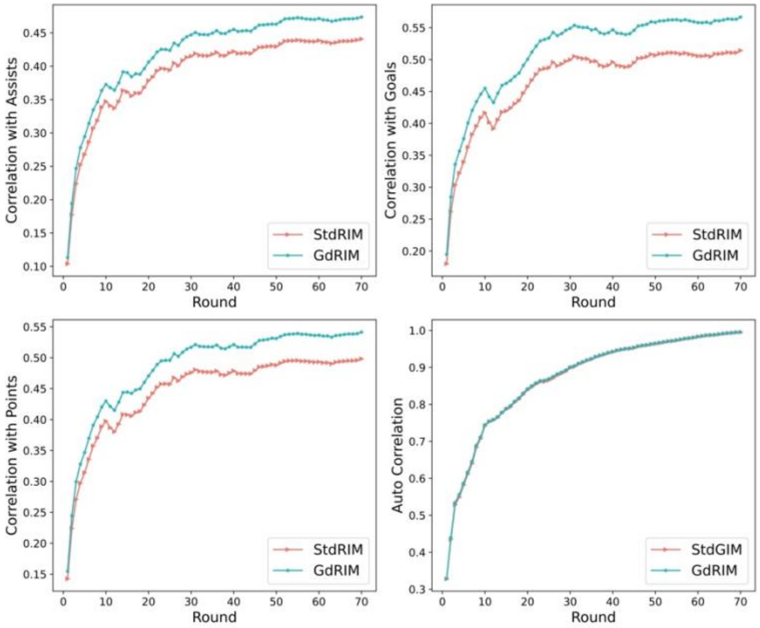

Figure 11.10: Round by round correlations between different player metrics and risk metrics. The figure plots correlations for the standard deviation and Gini deviation risk metrics (StdRIM and GdRIM).

Figure 11.11 shows the correlations between value-at-risk metrics and goal-related statistics. The correlations are even higher than with the deviation metrics (0.86 vs. 0.56 with goals). As the auto-correlation figure shows, a player's risk-taking as measured by value-at-risk is stable throughout a season. Our observations for team and player rankings therefore point in different directions: for teams, we have strong evidence that standard/Gini deviation measures risk-taking that indicates team strength, whereas for players, value-at-risk with a low confidence level seems to correlate better. Correlations with goal-based statistics are only a superficial signal of player strength, as goals occur rarely and cover only a small part of relevant actions. Further research into risk-taking by players seems warranted.

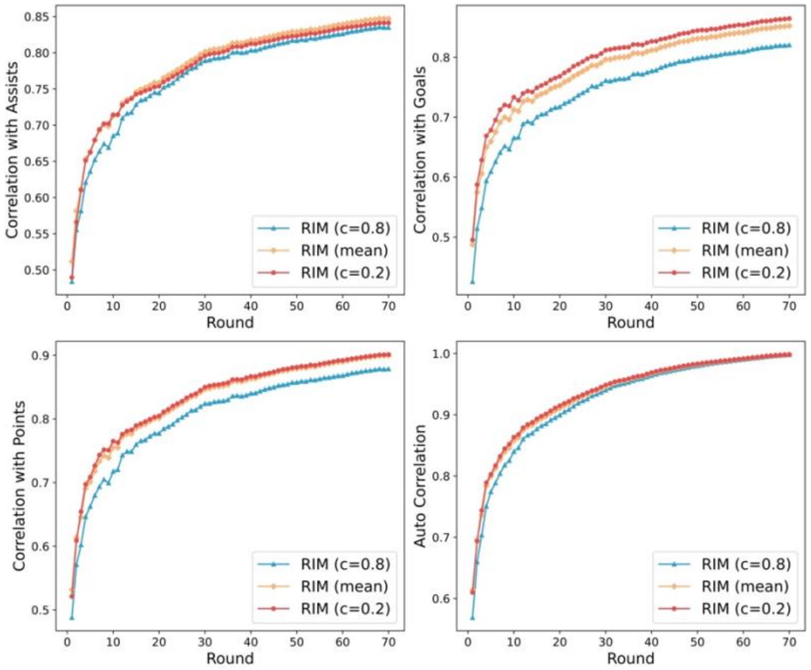

Figure 11.11: Round by round correlations between different player metrics and risk metrics. The figure plots correlations for value-at-risk metrics (RIM(c)), and goal impact metric (based on expected value). We plot risk-seeking confidence values (0.2), risk-averse confidence (0.8), and a neutral value with the confidence level corresponding set at the mean of the outcome distribution (its Q-value).

A possible explanation for why the deviation-based metrics correlate less with goals, and do worse by the eye test, is to distinguish two ways in which a player's actions can increase outcome variability: (1) *Deliberate Risks*, and (2) *Unforced Errors*. Strong players take controlled risks, based on confidence in their skills. For example, in ice hockey, a strong player will often carry the puck into their offensive zone, drawing defenders to themselves, which increases the risk of losing the puck but also increases the chance of a successful attack. In contrast, dumping the puck behind the defending team's goal is a safer move, but tends to lead to fewer goals (Schulte et al., 2017a). An analogue in soccer would be dribbling the ball towards the defenders' goal rather than passing it to a teammate. An example of an unforced error would be losing the ball during a promising attack, which causes the range of likely outcomes to spread from a concentration on a successful attack to both teams being likely to score. For a simple statistical model of this distinction, consider a Bernoulli model for a binary event with probability p . For example, an expected goals model might assign a probability p to a shot succeeding at time t .The variance of success is given by $p(1-p)$ and is maximal at p = 50%. A deliberate risk may increase the scoring chance from p < 50% towards $p' > p$ where p' < 50%; for example, a deliberate risk may increase the scoring chance from 30% to 40%. *Such a move increases both scoring chance and variance.* On the other hand, an error may move the scoring chance from p towards $p' < p$ with p' < 50%, for example from 70% to 60%. Such a move *decreases the scoring chance and increases variance*. It is possible that strong players' risks tend to be mainly of the beneficial deliberate type, whereas weaker players' risks are of the harmful error type, so both may display a high risk measure. This analysis suggests that a fruitful direction for future research is a player metric that combines risk and reward, e.g., the standard deviation and the mean of the outcome distribution associated with a player's action.

11.8 Conclusion

Decision-makers in sports often face a *trade-off between risk and reward*. Should a basketball player take a long-distance 3-point shot or a safer 2-point shot? Should a hockey player carry the puck, pass it to a teammate, or dump it behind the net? Studying the behavior of athletes and coaches when faced with risk-reward trade-offs requires tools for *risk analytics*. This paper described computational tools for risk analytics in sports by leveraging concepts and techniques from *distributional reinforcement learning*. Distributional RL aims to model the distribution of possible future outcomes, whereas traditional RL estimates the expected value. For representing the complex multi-modal outcome distributions that stem from sports dynamics, we adapted a state-of-the-art approach to distributional RL, which utilizes quantile regression as an expressive non-parametric framework for modelling distributions. We applied distribution RL techniques based on the Bellman equation to estimate a dynamic outcome distribution in the National Hockey League, for 1000+ games and 1M+ events. The literature on risk analysis has proposed different ways to quantify the risk inherent in an outcome distribution. We evaluated several of the most prominent ones for applications in hockey analytics, to answer the questions: Do stronger teams take more risks? Do stronger players take more risks? We found that the traditional standard deviation risk metric is an excellent predictor of team success: Both the standard deviation and Gini deviation of a team's outcome distribution, aggregated over a season, show a 0.90 correlation with the team's league standing at the end of a season (determined by their total points). Value-at-risk with confidence level 0.2 shows a lesser but still strong correlation of 0.51.

For player ranking, we found that value-at-risk with risk-seeking confidence level 0.2, aggregated over all actions by a player in a season, shows a very high correlation of 0.86 with the player's total season goals. Standard deviation and Gini deviation, in contrast, show slightly lower but still strong correlations of 0.51 and 0.56, respectively. We suggest that modelling the risk-taking behavior of players may require a more fine-grained metric that distinguishes between deliberate risks, which are incurred by actions requiring high skill, and risks stemming from errors (e.g., losing possession of the puck). A promising source of such fine-grained metrics are measures from portfolio theory that combine risk and reward (expected outcome). For example, we might investigate in sport analytics a version of the famous Sharpe ratio that divides expected value by standard deviation.

In sum, risk analytics is a promising new approach to sports analytics that focuses on the difficult trade-offs between taking risks and maximizing the chance of success that decision-makers in sports face. Distributional reinforcement learning provides the computational tools for estimating both risks and expected rewards in large sports data sets. Our NHL study shows that strong teams take big risks, and confirms to a lesser degree, that strong players are risk-takers as well.

11.9. Acknowledgements

This research was supported by discovery grants from Canada's Natural Sciences and Engineering Research Council, awarded to Schulte and Poupart. Yudong Luo is also supported by a David R. Cheriton Graduate Scholarship, and a Waterloo AI Institute Graduate Scholarship. Resources used in this work were provided by the Province of Ontario and the Government of Canada through CIFAR and companies sponsoring the Vector Institute.[5] The dataset analyzed in this paper was generously provided by Sportlogiq.

[5] [5]https://vectorinstitute.ai/partners/

11.10 Appendix

11.10.1 Learning Value Functions and Value Distributions

Figure 11.12 illustrates our recurrent neural network architecture for learning Q-values and distributional quantiles. Given a demonstration dataset of observed trajectories (i.e., $\mathcal{D} = \{\tau\}$), a Q-function satisfying the Bellman equation can be learned by minimizing the 2-Norm of the Temporal difference (TD) error, which is defined by:

$$\mathcal{L}(\theta) = \mathbb{E}_{\tau\sim\mathcal{D}}\left(\hat{Q}_k(s_{t+1}, a_{t+1}) + r_{k,t} - \hat{Q}_k(s_t, a_t)\right)^2 \qquad (11)$$

where the expectation represents the average over all transitions $(s_t, a_t; s_{t+1}, a_{t+1})$ observed in the dataset. The term $\widehat{Q_k}(s_t, a_t)$ represents the current value estimate, and $\widehat{Q_k}(s_{t+1}, a_{t+1}) + r_{k,t}$ a look-ahead step. Minimizing their squared difference drives the neural network to satisfy the Bellman equation (Equation (3)). For more details on implementing the temporal Bellman equation with quantile regression, please see Liu et al. (2022).

11.10.2 Action Impact on Value-at-Risk

Figure 11.13 shows the action impacts for Var(0.2) with confidence level 0.2 Game- changing events have a high impact on risk, as we observed in Section 5. In general the changes in risk are less than with standard/Gini deviations as this metric focuses on low probability outcomes.

Figure 11.14 shows the impact of actions on risk as measured by Var(0.8) with confidence level 0.8. At this confidence level, risk impact is similar to goal impact, that is, tends to measure increase in goal scoring chances rather than the increase in the variability of goal-scoring chances.

11.10.3 Top 20 Player Tables for other Risk Metrics

Table 11.8 shows the top 20 players for the Gini impact risk metric, which is very similar to the top 20 for the standard deviation risk metric.

As shown in Table 11.9, for the risk-averse confidence level 0.8, value-at-risk identifies many stars, such as Scheifele and Crosby. This metric shows a bias towards centres.

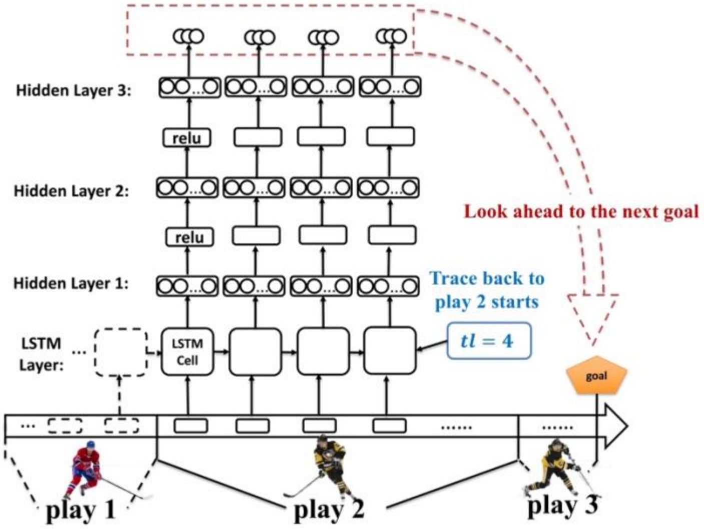

Figure 11.12: Our recurrent neural network architecture for learning Q-values and a distribution of action outcomes. At each time step, the RNN receives as input a pair s_t, a_t where the state s_t is a vector of features shown in Table 1. It outputs an estimate of the Q-value (chance of scoring the next goal) or a set of quantiles representing the distribution of action outcomes (goals scored). We utilize an LSTM architecture. For more details please see Liu et al. (2022).

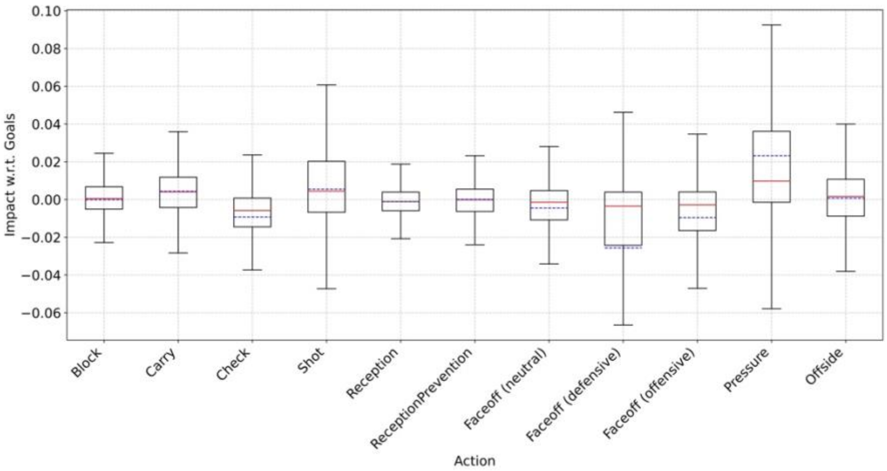

Figure 11.13: Box Plot with confidence 0.2 (risk-seeking) The red line indicates the median value and the blue line indicates the mean value. Each point represents the risk impact of the given action in a state.

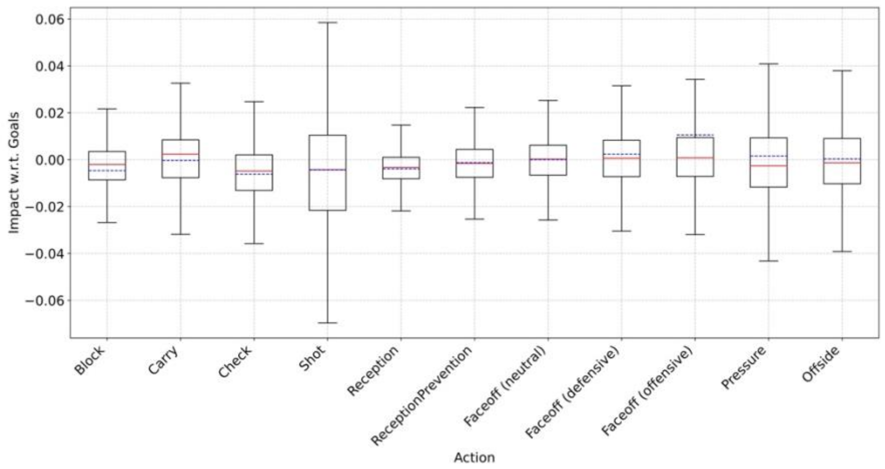

Figure 11.14: Box Plot with for RIM with confidence 0.8 (risk-averse) The red line indicates the median value and the blue line indicates the mean value. Each point represents the risk impact of the given action in a state.

Table 11.8: Top 20 players with GdRIM based on the entire season.

Player Name	Position	Team	P	A	G	GdRIM
Johnny Gaudreau	LW	CGY	99	63	36	7.15
Patrick Kane	RW	CHI	110	66	44	6.61
Nikita Kucherov	RW	TBL	128	87	41	6.53
Cam Atkinson	RW	CBJ	69	28	41	6.06
Alex Ovechkin	LW	WSH	89	38	51	6.00
Mitchell Marner	RW	TOR	94	68	26	5.98
Timo Meier	RW	SJS	66	36	30	5.97
Matthew Tkachuk	LW	CGY	77	43	34	5.18
Brad Marchand	LW	BOS	100	64	36	5.06
Kyle Connor	LW	WPG	66	32	34	5.03
Vladimir Tarasenko	RW	STL	68	35	33	4.96
Brendan Gallagher	RW	MTL	52	19	33	4.91
David Pastrnak	RW	BOS	81	43	38	4.87
Jaden Schwartz	LW	STL	36	25	11	4.87
Filip Forsberg	LW	NSH	50	22	28	4.64
Mikko Rantanen	RW	COL	87	56	31	4.62
Josh Anderson	RW	CBJ	47	20	27	4.52
Evgenii Dadonov	RW	FLA	70	42	28	4.44
Jason Zucker	LW	MIN	42	21	21	4.27
Jonathan Huberdeau	LW	FLA	92	62	30	4.22

Table 11.9: Top 20 players with RIM at risk-averse confidence 0.8 based on the entire season.

Player Name	Position	Team	P	A	G	RIM(0.8)
Aleksander Barkov	C	FLA	96	61	35	50.50
Leon Draisaitl	C	EDM	105	55	50	49.67
Mark Scheifele	C	WPG	84	46	38	48.73
Sidney Crosby	C	PIT	100	65	35	47.24
Jonathan Toews	C	CHI	81	46	35	45.22
Mitchell Marner	RW	TOR	94	68	26	42.71

Dylan Larkin	C	DET	73	41	32	41.82
Nikita Kucherov	RW	TBL	128	87	41	41.00
Max Domi	LW	MTL	72	44	28	40.47
Connor McDavid	C	EDM	116	75	41	40.45
Bo Horvat	C	VAN	61	34	27	39.92
Mika Zibanejad	C	NYR	74	44	30	39.29
Artemi Panarin	LW	CBJ	87	59	28	38.93
Sebastian Aho	C	CAR	83	53	30	38.55
Mark Stone	RW	VGK	73	40	33	38.38
Claude Giroux	C	PHI	85	63	22	37.98
Johnny Gaudreau	LW	CGY	99	63	36	37.86
Mathew Barzal	C	NYI	62	44	18	37.83
Nicklas Backstrom	C	WSH	74	52	22	37.78
Brad Marchand	LW	BOS	100	64	36	37.55

References

Beaudoin, D., & Swartz, T. B. (2010). Strategies for pulling the goalie in hockey. The American Statistician, 64(3), 197-204.

Bellemare, M. G., Dabney, W., & Munos, R. (2017, July). A distributional perspective on reinforcement learning. In International conference on machine learning (pp. 449-458). PMLR.

Chan, T. C., Fernandes, C., & Puterman, M. L. (2021). Points gained in football: Using markov process-based value functions to assess team performance. Operations Research, 69(3), 877-894.

Decroos, T., Dzyuba, V., Haaren, J. V. and Davis, J. (2017), Predicting Soccer Highlights from Spatio-Temporal Match Event Streams, *in* 'AAAI 2017', pp. 1302--1308.

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019, July). Actions speak louder than goals: Valuing player actions in soccer. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 1851-1861).

Franks, A. M., D'Amour, A., Cervone, D., & Bornn, L. (2016). Meta-analytics: tools for understanding the statistical properties of sports metrics. Journal of Quantitative Analysis in Sports, 12(4), 151-165.

Littman, M. L. (1994). Markov games as a framework for multi-agent reinforcement learning. In Machine learning proceedings 1994 (pp. 157-163). Morgan Kaufmann.

Liu, G., & Schulte, O. (2018). Deep reinforcement learning in ice hockey for context-aware player evaluation. arXiv preprint arXiv:1805.11088.

Liu, G., Luo, Y., Schulte, O., & Kharrat, T. (2020). Deep soccer analytics: learning an action-value function for evaluating soccer players. Data Mining and Knowledge Discovery, 34, 1531-1559.

Liu, G., Luo, Y., Schulte, O., & Poupart, P. (2022). Uncertainty-aware reinforcement learning for risk-sensitive player evaluation in sports game. Advances in Neural Information Processing Systems, 35, 20218-20231.

Luo, Y., Schulte, O., & Poupart, P. (2020, July). Inverse reinforcement learning for team sports: Valuing actions and players. In *IJCAI* (pp. 3356-3363).

Luo, Y., Liu, G., Poupart, P., & Pan, Y. (2023). An alternative to variance: Gini deviation for risk-averse policy gradient. *Advances in Neural Information Processing Systems*, *36*, 60922-60946.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *nature*, *518*(7540), 529-533.

National Hockey League. National hockey league official rules 2014-2015. http://www.nhl.com/nhl/en/v3/ext/rules/2014-2015-rulebook.pdf.

Pelechrinis, K. (2016). Decision Making in American Football: Evidence from 7 Years of NFL Data. In *MLSA@ PKDD/ECML*.

Pettigrew, S. (2015). Assessing the offensive productivity of NHL players using in-game win probabilities. In *9th annual MIT sloan sports analytics conference* (Vol. 2, No. 3, p. 8).

Routley, K. D. (2015). A markov game model for valuing player actions in ice hockey.

Schulte, O. (2022, September). Valuing Actions and Ranking Hockey Players With Machine Learning. In *Linköping Hockey Analytics Conference* (pp. 2-9).

Schulte, O., Khademi, M., Gholami, S., Zhao, Z., Javan, M., & Desaulniers, P. (2017). A Markov Game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery*, *31*, 1735-1757.

Schulte, O., Zhao, Z., Javan, M., & Desaulniers, P. (2017, March). Apples-to-apples: Clustering and ranking NHL players using location information and scoring impact. In *Proceedings of the MIT Sloan Sports Analytics Conference*.

Sutton, R. S., & Barto, A. G. (1998). The reinforcement learning problem. *Reinforcement learning: An introduction*, 51-85.

Information about the authors

Oliver Schulte (corresponding author)
School of Computer Science
Simon Fraser University, Vancouver
oschulte@cs.sfu.ca

Sheng Xu
School of Data Science,
The Chinese University of Hong Kong, Shenzhen
Shenzhen, China
shengxu1@link.cuhk.edu.cn

Yudong Luo
Cheriton School of Computer Science
University of Waterloo
Waterloo, Canada
Vector Institute, Toronto, Canada
yudong.luo@uwaterloo.ca

Pascal Poupart
Cheriton School of Computer Science
University of Waterloo
Waterloo, Canada
Vector Institute, Toronto, Canada
ppoupart@uwaterloo.ca

Guiliang Liu

School of Data Science,
The Chinese University of Hong Kong, Shenzhen
Shenzhen, China
liuguiliang@cuhk.edu.cn