Distributional State Aggregation in Reinforcement Learning

August XX, 2020

Shayan Amani

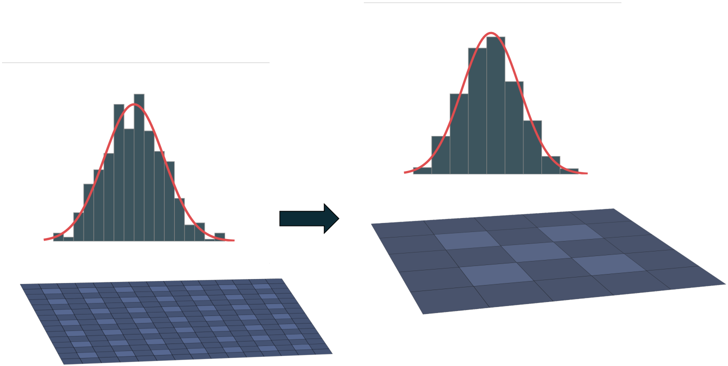
Department of Computer Science, University of New Hampshire

whoami

- Finished bachelor's in EE in spring 2015
- Aspired to develop for my own ideas
  - at a certain point, I had 200k monthly active users
- Started Ph.D. at UNH in fall 2017
- Have been experimenting with quite a few areas to discover what I like to pursue

Introduction

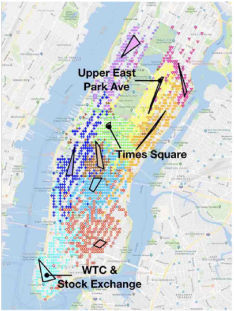
Problem



- Reduce model into low resolution clusters.
- Same distribution, different discritization in state space.

- No general rule applicable to various aggregation problems.
- Domain-dependent solutions.
- feature-based aggregation and more engineering.

Example



- 11 million taxi trips in Manhattan, NYC.
- Clustered into 1922 divisions.

Yaqi Duan, Zheng Tracy Ke, and Mengdi Wang. "State Aggregation Learning from Markov Transition Data". In: (2019), pp. 4486–4495. arXiv: 1811.02619. URL: http://papers.nips.cc/paper/8698-state-aggregation-learning-from-markov-transition-data.pdf.

- less computational power
- more tractable problem
- analytically transparent approximation
  - compared to Neural Networks

Steps toward an aggregation framework:

- Non-parametric
- Domain-agnostic
- Sample-based

Outline

Density Estimation

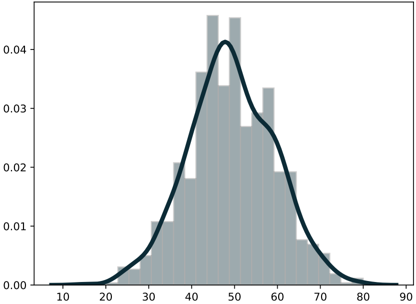
Density Estimators

Definition:

Density estimation is fitting an estimate, based on **observed data** of an unobservable underlying **probability density function (PDF)**.

Estimators:

- histogram
- kernel density estimation
- wavelets thresholding
- smoothing splines



$$\hat{f}(x; w) = \frac{|B_j|}{nw}, \quad x \in B_j$$

Histograms

Advantages:

- As density estimators, histograms have been studied thoroughly for decades.
- Computational advantages compared to kernel-based methods.
- Non-parametric estimation to avoid domain-specific feature engineering.

Drawbacks:

- No universal optimality conditions on parameters (k, w) and asymptotic considerations.
- Cost of being non-parametric: slow convergence rate.
- Strong theoretical assumptions render theorems and results impractical.
- Computational complexity in case of non-normal distributions.

State aggregation based on histograms:

- Automatic and efficient method of choosing the number of bins.
- Based on the characteristics of the underlying distribution.

Metrics

- **risk** or integrated mean squared error (**IMSE**).
  - No direct solution to IMSE, underlying distribution is unknown
- **estimated risk** or cross-validation estimator of risk.

IMSE
Measure of Fit

$$\begin{aligned}
\mathrm{MSE}\{\hat{f}(.;w)\} &= \mathbb{E}\left[\hat{f}(x) - f(x)\right]^2 \\
&= \frac{1}{wk}\hat{f}(x) - \frac{1}{k}\hat{f}(x)^2 + \left[\hat{f}(x) - f(x)\right]^2 \qquad (1) \\
&= \text{Variance} + \text{Bias}^2
\end{aligned}$$

- The histogram converges in mean squared to $f(x)$ if $w \to 0$ and $nw \to \infty$.

$$\mathrm{IMSE}\{\hat{f}(.;w)\} = \int \mathbb{E}\{\hat{f}(x) - f(x)\}^2 dx \tag{2}$$

- Global error measure of a histogram estimate.
- Slower convergence to a fixed-point than parametric estimators[2]: $n^{-2/3} > n^{-1}$

[2] Larry Wasserman. *All of Statistics; A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer New York, 2004. DOI: 10.1007/978-0-387-21736-9. URL: http://link.springer.com/10.1007/978-0-387-21736-9.

Metrics

How many bins?

Bias-Variance Tradeoff

Number of buckets of discritization:

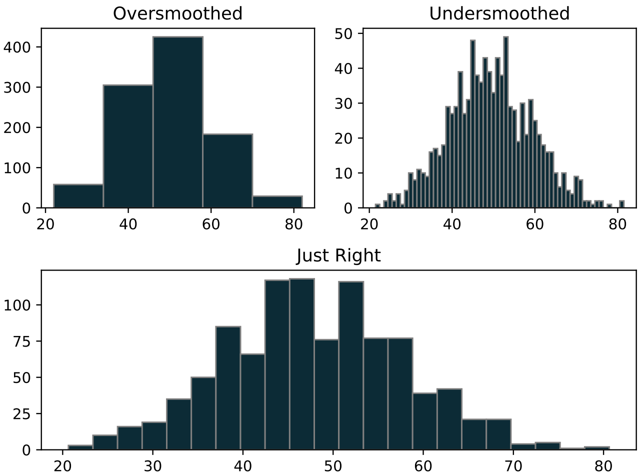
- range split aggregation

Variance **Bias**

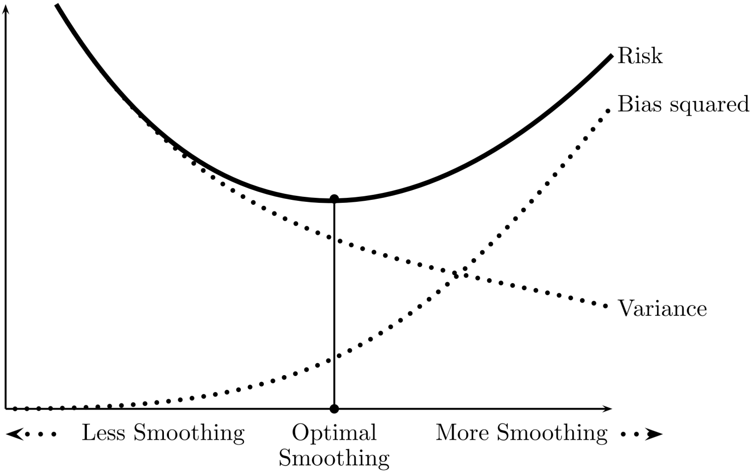
Bin-count **Bin-width**

$$\Longleftrightarrow$$

Smoothing
Bias-Variance Tradeoff



Optimal Smoothing
Bias-Variance Tradeoff



Wasserman, *All of Statistics; A Concise Course in Statistical Inference*.

Bin Count vs. Bin Width

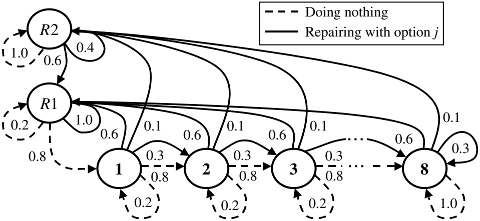
For equally-spaced intervals:

$$k = \left\lceil \frac{\max(x) - \min(x)}{w} \right\rceil \tag{3}$$

- depends on the distribution range and interval size: $k = f(R, w)$

Methods

Machine Replacement



Erick Delage and Shie Mannor. "Percentile optimization for Markov decision processes with parameter uncertainty". In: *Operations Research* 58.1 (2010), pp. 203–213. ISSN: 0030364X. DOI: 10.1287/opre.1080.0685.

Methods

Interval Count Methods

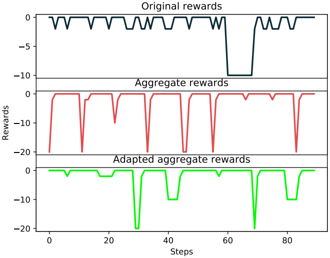
$$w = \frac{\max(x) - \min(x)}{\sqrt{n}} \tag{4}$$

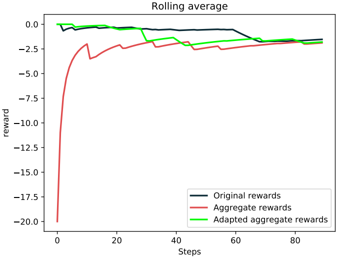
$$k = \lceil \sqrt{n} \rceil \tag{5}$$

- Intuitive: no restricting assumptions
- Used by Excel[5].

[5] *EXCEL Univariate: Histogram.* URL: http://cameron.econ.ucdavis.edu/excel/ex11histogram.html.

Aggregation
Square-root Rule





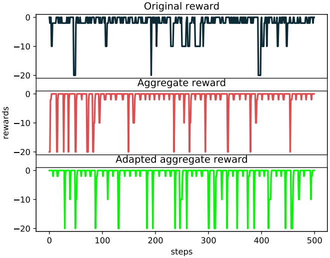
$$w = \frac{\max(x) - \min(x)}{1 + \lg n} \tag{6}$$

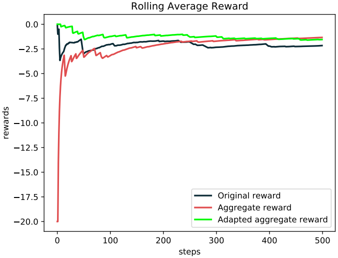
$$k = 1 + \lceil \lg n \rceil \tag{7}$$

- Estimates the original distribution with a series of binomial coefficients[6].
- Normal distribution is implied.

[6] Herbert A. Sturges. "The Choice of a Class Interval". In: *Journal of the American Statistical Association* 21.153 (1926), pp. 65–66. URL: http://www.jstor.org/stable/2965501.

Aggregation
Sturge's Formula





Rice Rule
Interval Count

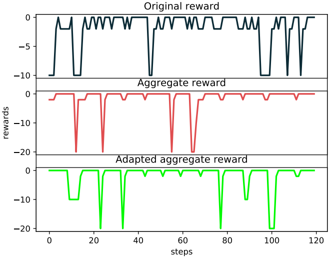
$$w = \frac{\max(x) - \min(x)}{2\sqrt[3]{n}} \tag{8}$$

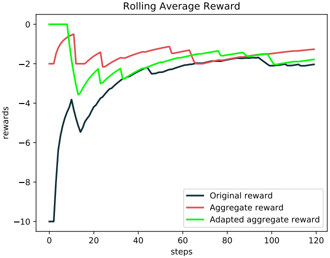
$$k = \lceil 2\sqrt[3]{n} \rceil \tag{9}$$

- The intuitive successor to Sturges' rule.
- Recommended by trial and errors[7].

[7] David Lane. "Online Statistics Education: A Multimedia Course of Study". In: *EdMedia + Innovate Learning* 2003.1 (2003), pp. 1317–1320.

Aggregation
Rice Rule





Doane's Rule
Interval Count

$$k = \underbrace{1 + \lceil \lg n \rceil}_{\text{Sturges' classes}} + \underbrace{\left\lceil \lg \left(1 + \frac{\left|\sqrt{\beta_1}\right|}{\sigma_{\sqrt{\beta_1}}} \right) \right\rceil}_{\text{Doane's extended classes}}$$

$$\sqrt{\beta_1} = \frac{m_3}{m_2^{3/2}} \tag{10}$$

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2, m_3 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^3 \tag{11}$$

$$\sigma_{\sqrt{\beta_1}} = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} \tag{12}$$

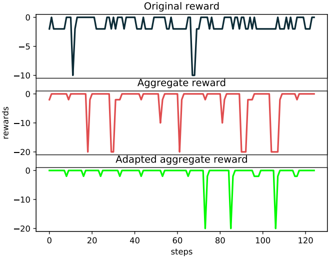
Doane's Formula
Interval Count

- Modified Sturges' formula to reflect distribution characteristics[8].
- Hence, performs better in case of skewed distributions[9].
- Inspired by coding in the information theory: increasing entropy by introducing more symbols for coding.

[8] David P Doane. *Aesthetic Frequency Classifications*. Tech. rep. 4. 1976, pp. 181–183.

[9] David P Doane and Lori E Seward. *Measuring Skewness: A Forgotten Statistic?*. Tech. rep. 2. 2011. URL: www.amstat.org/publications/jse/v19n2/doane.pdf.

Aggregation
Doane's Formula



Methods

Interval Width Methods

$$\text{IMSE} = \int E\{\hat{f}(x) - f(x)\}^2 dx \tag{13}$$

using Taylor expansion:

$$\text{IMSE} = 1/\left(nw\right) + \frac{1}{12} w^2 \int_{-\infty}^{\infty} f'(x)^2 dx + O\left(1/n + w^3\right) \tag{14}$$

$$w^* = \left\{ 6 \Big/ \int_{-\infty}^{\infty} f'(x)^2 dx \right\}^{1/3} n^{-1/3} \tag{15}$$

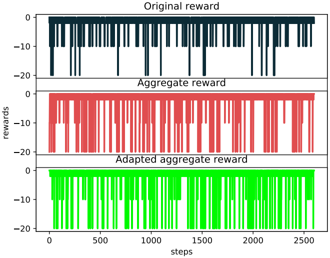
$$\begin{aligned} w^* &= 2\sqrt[6]{9\pi}\,\sigma\, n^{-1/3} \\ &= 3.49083\,\sigma\, n^{-1/3} \end{aligned} \tag{16}$$

Scott's Formula
Interval Width

- Calculates the w^* by minimizing IMSE[10].
- Derived the estimate for a Gaussian distribution.

[10] David W Scott. "On optimal and data-based histograms". In: *Source: Biometrika* 66.3 (1979), pp. 605–610.

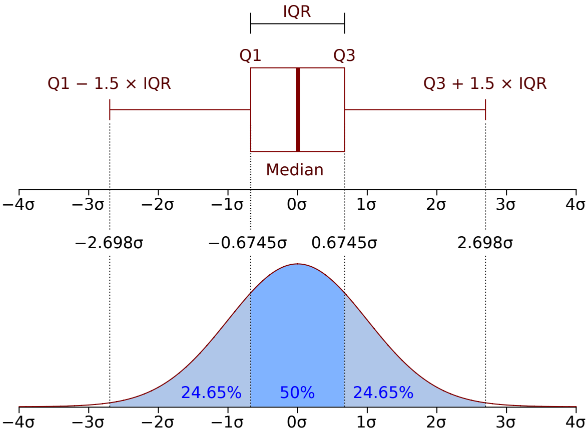
Aggregation
Scott's Formula



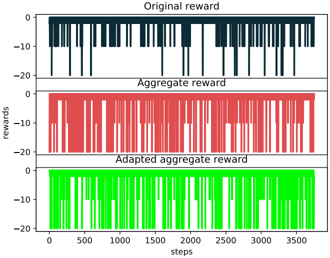
$$w = 2\,\mathrm{IQR}(x)\,n^{-1/3} \tag{17}$$

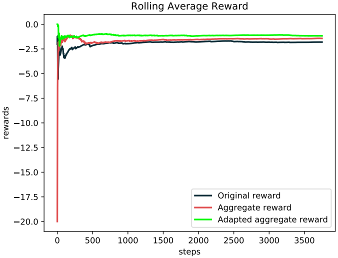
- $3.5\,\sigma \approx 2\,\mathrm{IQR}$
- More robustness to outliers than standard deviation

IQR
Freedman-Diaconis' Rule



Aggregation
Freedman-Diaconis' Rule





Shimazaki-Shinomoto's Choice
Interval Width

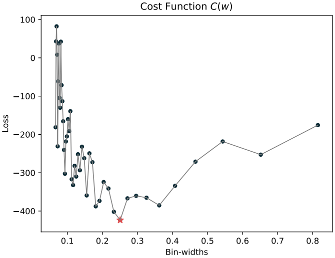
$$w^* = \arg\min_w C(w) = \arg\min_w \frac{2\bar{X} - \sigma^2}{w^2} \tag{18}$$

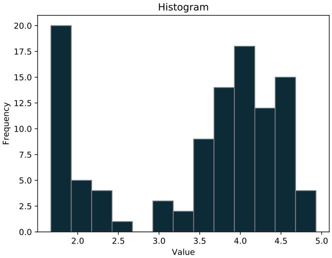
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i;\ \underbrace{\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}_{\text{biased variance}}$$

- Minimizes cost function for a set of proposed bin-widths[12].
- Looks over the dispersion of data points count falling into bins.

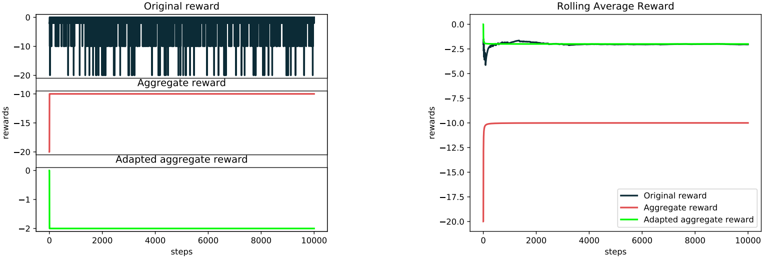
[12] Hideaki Shimazaki and Shigeru Shinomoto. "A recipe for optimizing a time-histogram". In: *Advances in Neural Information Processing Systems*. 2007, pp. 1289–1296. ISBN: 9780262195683.

Cost Analysis
Shimazaki-Shinomoto's Choice





Aggregation
Shimazaki-Shinomoto's Choice



Discussion

- Incorporating more characteristics of underlying distribution leads to a better aggregation policy.
- High dimensional spaces may take more advantage of this approach.
- Sampling distribution decently explains the candidacy of states for aggregation.

Ideas for Future

1. Adjusted Fisher-Pearson estimation of skewness to take sample size into account.
2. Employ studentized test instead of kurtosis in Doane's formula[13],[14].
3. Variable bin-width aggregation.

[13] R C Geary. *Moments of the Ratio of the Mean Deviation to the Standard Deviation for Normal Samples*. Tech. rep. 3. 1936, pp. 295–307.

[14] Ronald L. Tracy and David P. Doane. "Using The Studentized Range to Assess Kurtosis". In: *Journal of Applied Statistics* 32.3 (2005), pp. 271–280. ISSN: 02664763. DOI: 10.1080/02664760500054632.

Questions

"Our virtues and our failings are **inseparable**, like force and matter. **When they separate, man is no more**."

Nikola Tesla

Duan, Yaqi, Zheng Tracy Ke, and Mengdi Wang. “State Aggregation Learning from Markov Transition Data”. In: (2019), pp. 4486–4495. arXiv: 1811.02619. URL: http://papers.nips.cc/paper/8698-state-aggregation-learning-from-markov-transition-data.pdf.

Wasserman, Larry. *All of Statistics; A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer New York, 2004. DOI: 10.1007/978-0-387-21736-9. URL: http://link.springer.com/10.1007/978-0-387-21736-9.

Delage, Erick and Shie Mannor. “Percentile optimization for Markov decision processes with parameter uncertainty”. In: *Operations Research* 58.1 (2010), pp. 203–213. ISSN: 0030364X. DOI: 10.1287/opre.1080.0685.

EXCEL Univariate: Histogram. URL: http://cameron.econ.ucdavis.edu/excel/ex11histogram.html.

Sturges, Herbert A. “The Choice of a Class Interval”. In: *Journal of the American Statistical Association* 21.153 (1926), pp. 65–66. URL: http://www.jstor.org/stable/2965501.

- Lane, David. "Online Statistics Education: A Multimedia Course of Study". In: *EdMedia + Innovate Learning* 2003.1 (2003), pp. 1317–1320.
- Doane, David P. *Aesthetic Frequency Classifications*. Tech. rep. 4. 1976, pp. 181–183.
- Doane, David P and Lori E Seward. *Measuring Skewness: A Forgotten Statistic?*. Tech. rep. 2. 2011. URL: www.amstat.org/publications/jse/v19n2/doane.pdf.
- Scott, David W. "On optimal and data-based histograms". In: *Source: Biometrika* 66.3 (1979), pp. 605–610.
- Shimazaki, Hideaki and Shigeru Shinomoto. "A recipe for optimizing a time-histogram". In: *Advances in Neural Information Processing Systems*. 2007, pp. 1289–1296. ISBN: 9780262195683.
- Geary, R C. *Moments of the Ratio of the Mean Deviation to the Standard Deviation for Normal Samples*. Tech. rep. 3. 1936, pp. 295–307.

Tracy, Ronald L. and David P. Doane. "Using The Studentized Range to Assess Kurtosis". In: *Journal of Applied Statistics* 32.3 (2005), pp. 271–280. ISSN: 02664763. DOI: 10.1080/02664760500054632.