Week 7: CVaR in Reinforcement Learning (Formulasi Rockafellar–Uryasev)

A

September 29, 2025

Agenda Sesi

Motivasi: Mengelola Tail Risk

- Minggu 6 : Distributional RL (QR-DQN) $\Rightarrow$ melihat *sebaran* return, bukan hanya rata-rata.
- Pertanyaan: bagaimana jika agent ingin **menghindari kerugian ekstrem**, bukan sekadar memaksimalkan $\mathbb{E}[R]$?
- Aplikasi: **asuransi** (klaim catastropic), **portofolio** (crash pasar), **operasi** (rare failures).
- Solusi: gunakan **ukuran risiko koheren**
- Fokus sesi ini: **CVaR**.

Recall: VaR dan Keterbatasannya

Value at Risk (quantile)

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} \mid \Pr(X \leq x) \geq \alpha\}$$

- Intuisi: kerugian ambang pada tingkat kepercayaan α.
- Keterbatasan: tidak selalu *subadditive* $\Rightarrow$ *tidak koheren*.

Conditional Value at Risk (CVaR)

Definisi CVaR

$$\mathrm{CVaR}_\alpha(X) = \mathbb{E}[X \mid X \geq \mathrm{VaR}_\alpha(X)].$$

- Artinya: rata-rata kerugian di **atas ambang** (VaR).
- CVaR lebih informatif daripada VaR karena melihat **seluruh ekor**.
- Cocok untuk mengukur risiko katastrofik.

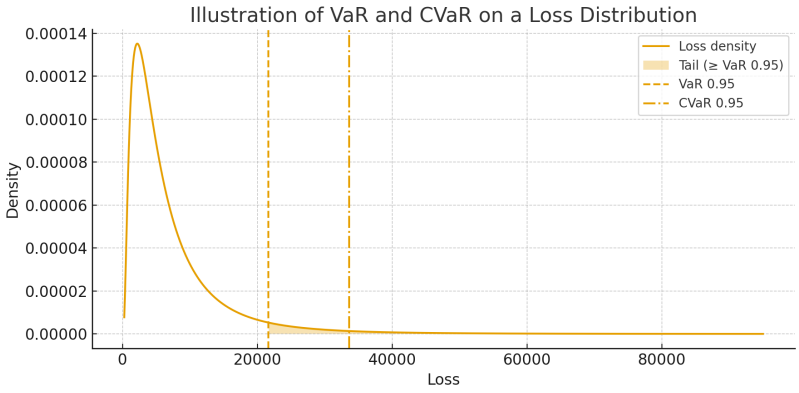
Formulasi Rockafellar–Uryasev (2000)

Representasi Optimisasi

$$\mathrm{CVaR}_\alpha(X) \;=\; \min_{\eta\in\mathbb{R}} \left\{ \eta + \frac{1}{1-\alpha}\,\mathbb{E}\big[(X-\eta)^+\big] \right\}, \qquad (z)^+ = \max(z,0).$$

- η berperan sebagai **ambang (threshold)**.
- Fungsi $(X-\eta)^+$ hanya menghitung **bagian ekor** di atas η.
- Minimum terjadi saat $\eta = \mathrm{VaR}_\alpha(X)$.

Intuisi Visual CVaR



- VaR_α: batas kuantil α (misalnya 95%).
- CVaR_α: rata-rata kerugian di area merah (ekor distribusi).
- Intuisi: CVaR melihat **seberapa parah kerugian** ketika sudah melewati VaR.

Mengubah Objective di RL

Risk-neutral RL

$$\max_{\pi_\theta} J(\theta) = \mathbb{E}_{\pi_\theta}[R]$$

- Agent memilih kebijakan π_θ untuk memaksimalkan rata-rata return.
- Semua hasil (baik atau buruk) hanya dihitung melalui nilai ekspektasi.
- Akibatnya : agent bisa memilih strategi yang memberikan rata-rata tinggi, meskipun ada kemungkinan kerugian besar (tail risk).
- Contoh: Dalam investasi, memilih saham yang rata-rata return-nya tinggi, meskipun sesekali bisa anjlok drastis.

CVaR-RL

$$\max_{\pi_\theta} J_\alpha(\theta) = \mathrm{CVaR}_\alpha(R)$$

- Agent memilih kebijakan π_θ untuk memaksimalkan CVaR dari return R.
- Artinya, agent fokus pada rata-rata hasil di bagian ekor terburuk (misalnya 5% skenario terburuk).
- Dengan demikian, strategi yang dipilih lebih konservatif dan stabil, karena meminimalkan risiko kerugian besar.
- Contoh : Dalam asuransi, perusahaan lebih peduli terhadap klaim katastrofik $\rightarrow$ sehingga premi ditentukan berdasarkan CVaR, bukan sekadar expected loss.

Adaptasi CVaR pada Berbagai Algoritma RL

Algoritma RL	Risk-neutral Objective	Adaptasi dengan CVaR
Policy Gradient (REINFORCE)	$J(\theta) = \mathbb{E}_{\pi_\theta}[R]$	Ganti dengan $J_\alpha(\theta) = \mathrm{CVaR}_\alpha(R)$. Update gradien pakai surrogate Rockafellar–Uryasev.
ActorCritic (A2C, PPO, SAC)	Actor memaksimalkan $\mathbb{E}[R]$, Critic mengestimasi $V^\pi(s)$	Actor memaksimalkan $\mathrm{CVaR}_\alpha(R)$. Critic dipakai untuk mengestimasi tail expectation.
Value-based (DQN)	$Q^\pi(s,a) = \mathbb{E}[R \mid s,a]$	Definisikan $Q^\pi_\alpha(s,a) = \mathrm{CVaR}_\alpha(R \mid s,a)$. Update Q berdasarkan tail losses.
Distributional RL (C51, QR-DQN)	Belajar distribusi return $Z^\pi(s,a)$	Estimasi quantile untuk level α, gunakan VaR_α dan CVaR_α untuk update/policy.
Model-based RL	Optimisasi $\mathbb{E}[R]$ dari rollout model	Optimisasi $\mathrm{CVaR}_\alpha(R)$ dari distribusi simulasi masa depan.

Inti: semua algoritma RL bisa dibuat risk-sensitive dengan mengganti objektif $\mathbb{E}[R]$ menjadi ukuran risiko (misal CVaR).

Policy Gradient: Risk-Neutral

Gradien standar

$$\nabla_\theta J(\theta) = \mathbb{E}_{\pi_\theta}[\nabla_\theta \log \pi_\theta(a|s)\ R]$$

- Ini adalah policy gradient biasa (REINFORCE).
- Artinya: parameter kebijakan θ diperbarui untuk memaksimalkan expected return.
- Tidak ada kontrol khusus terhadap risiko $\rightarrow$ bisa pilih strategi dengan return rata-rata tinggi, walaupun punya ekor buruk.

Policy Gradient: CVaR dengan Surrogate Function (Rockafellar–Uryasev)

Definisi Surrogate

$$L_\alpha(\theta, \eta) \;=\; \eta + \frac{1}{1-\alpha}\,\mathbb{E}_{\pi_\theta}\big[(R-\eta)^+\big], \qquad (z)^+ = \max(z, 0).$$

- **Surrogate** = fungsi bantu/pendekatan yang lebih mudah dihitung/dioptimasi dibanding definisi asli.
- Definisi CVaR:

$$\mathrm{CVaR}_\alpha(R) = \min_\eta L_\alpha(\theta, \eta).$$

- Jadi $L_\alpha(\theta, \eta)$ dipakai sebagai **fungsi objektif alternatif** (surrogate objective).
- Optimisasi dilakukan terhadap θ (policy) sekaligus η (estimasi VaR).

Gradien Surrogate

$$\nabla_\theta L_\alpha(\theta, \eta) = \frac{1}{1-\alpha}\, \mathbb{E}_{\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a|s)\,(R-\eta)^+\right]$$

$$\Rightarrow \text{ update policy fokus pada reward di ekor}$$

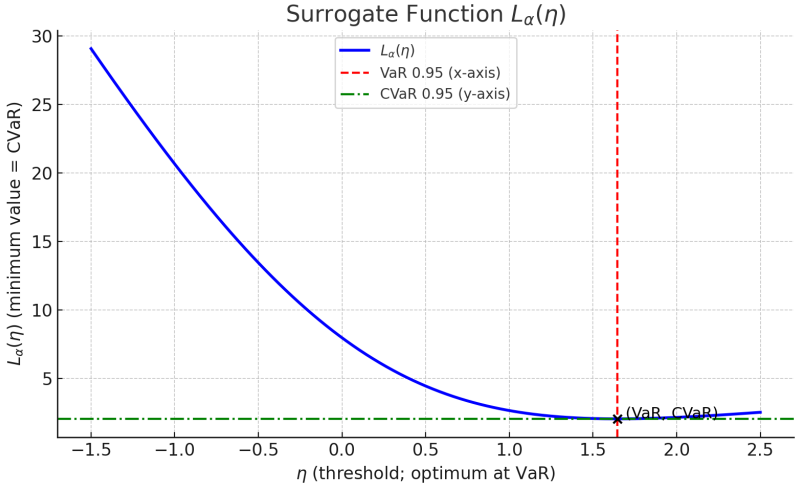
$$\nabla_\eta L_\alpha(\theta, \eta) = 1 - \frac{1}{1-\alpha}\, \Pr_{\pi_\theta}(R \geq \eta)$$

$$\Rightarrow \text{ update } \eta \text{ agar mendekati } \mathrm{VaR}_\alpha(R)$$

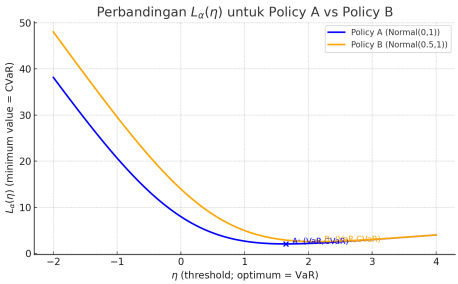
- **Peran** η:
  - η digeser-geser untuk mencari titik minimum.
  - Kondisi $\nabla_\eta L = 0 \;\Rightarrow\; \eta^* = \mathrm{VaR}_\alpha(R)$.
- **Peran** θ:
  - Saat $\eta = \eta^*$, nilai $L = \mathrm{CVaR}_\alpha(R)$.
  - Gradien $\nabla_\theta L_\alpha(\theta, \eta^*)$ = gradien $\mathrm{CVaR}_\alpha(R)$.
- **Intinya**:
  - Update $\eta \;\Rightarrow$ memastikan L benar-benar sama dengan CVaR.
  - Update $\theta \;\Rightarrow$ memaksimalkan CVaR lewat gradien surrogate.

Identitas dasar

$$\mathrm{CVaR}_\alpha(R) \;=\; \min_{\eta\in\mathbb{R}} L_\alpha(\theta,\eta), \qquad L_\alpha(\theta,\eta) = \eta + \frac{1}{1-\alpha}\,\mathbb{E}[(R-\eta)^+].$$



Pengaruh Parameter Policy (θ) terhadap $L_\alpha(\eta)$



- Optimisasi **terhadap** η: menemukan VaR untuk policy tertentu.
- Optimisasi **terhadap** θ: mengubah distribusi $\Rightarrow$ menggeser kurva agar CVaR lebih besar.

Mini-Project: CliffWalking (Tail Risk)

- Bandingkan dua agent: **REINFORCE (risk-neutral)** vs **CVaR-Policy-Gradient (risk-averse)**.
- Lingkungan: *CliffWalking* tabular (softmax policy; update episodik).
- Metrik evaluasi:
  - Mean return per episode,
  - lower-tail $\mathrm{VaR}_{0.05}$ dan $\mathrm{CVaR}_{0.05}$ (untuk return buruk),
  - frekuensi jatuh ke jurang.
- Ekspektasi hasil: **CVaR-PG** memilih rute lebih aman (rata-rata return sedikit lebih rendah, namun **tail risk lebih kecil** & jarang jatuh).

Studi Data: Insurance Pricing (Tail Risk in Industry)

- Data klaim: simulasi *Lognormal + Pareto mixture* (atau ganti dengan Danish Fire/CAS Auto Claims).
- Estimasi ukuran risiko pada level $\alpha \in \{0.95, 0.99\}$: VaR_α dan CVaR_α.
- Bandingkan premi:
  - **Expected-Loss Premium** $= \mathbb{E}[\text{loss}]$,
  - **CVaR-based Premium** $= \mathrm{CVaR}_\alpha(\text{loss})$ (lebih konservatif terhadap klaim ekstrem).
- Visualisasi: histogram losses + garis VaR_α dan CVaR_α (skala-y log untuk ekor).
- Diskusi: dampak pada **kestabilan modal** dan **proteksi katastrofik**.

Catatan implementasi

Jika memakai data asli, ganti bagian simulasi dengan `pd.read_csv(...)`; opsional: *tail modeling* (mis. GPD) untuk estimasi ekor.

Trade-off: Expected Gain vs Stability

- Risk-neutral: gain rata-rata tinggi, tail risk tinggi.
- Sensitif parameterisasi α (semakin dekat 1 $\Rightarrow$ makin konservatif).
- CVaR-RL: gain rata-rata sedikit turun, **stabilitas meningkat**.
- Cocok untuk asuransi & investor konservatif.

Diskusi kelas

Kapan organisasi sebaiknya memilih **CVaR** alih-alih risk-neutral? Kaitkan dengan regulasi (modal berbasis risiko).

Panduan Diskusi Kelas

CliffWalking (RL simulasi):

- Apa perbedaan perilaku agent **REINFORCE** dan **CVaR-PG**?
- Mengapa CVaR-PG lebih aman meski return rata-rata turun?
- Dalam konteks nyata (robotika, operasi industri), kapan strategi konservatif ini lebih diinginkan?

Insurance Pricing (data klaim):

- Apa implikasi perbedaan premi antara **expected-loss** vs **CVaR-based**?
- Bagaimana sensitivitas hasil terhadap level α (95% vs 99%)?
- Mengapa perusahaan asuransi cenderung memilih pendekatan CVaR dalam pricing atau modal cadangan?

Jembatan RL $\leftrightarrow$ Industri:

- Apa benang merah antara CliffWalking dan Insurance Pricing?
- Apakah trade-off *expected gain vs stability* muncul di kedua domain ini?
- Bagaimana konsep *tail risk* di RL bisa diterjemahkan ke risiko katastrofik di asuransi?

Referensi Singkat

- Rockafellar, R. T., & Uryasev, S. (2000). *Optimization of Conditional Value-at-Risk*. Journal of Risk.
- Tamar, A., Glassner, Y., & Mannor, S. (2015). *Optimizing the CVaR via Sampling*. AAAI.
- Chow, Y., et al. (2015–2017). *Risk-Sensitive and CVaR MDPs*.