

# Week 8: Entropic Value-at-Risk (EVaR)

## Advanced Risk Measures in Reinforcement Learning

Aprida Siska

Actuarial Computation and Simulation (ACS)

Semester I

# Agenda Sesi

- 1 Motivasi
- 2 Konsep EVaR (vs VaR/CVaR)
- 3 EVaR dalam RL
- 4 Policy Gradient untuk EVaR
- 5 Demo dan Implementasi
- 6 Analisis Trade-off
- 7 Referensi

# Motivasi: Mengelola Tail Risk Lebih Ketat

Minggu 7: CVaR-RL (Rockafellar–Uryasev)  $\Rightarrow$  fokus pada rata-rata ekor terburuk.

Pertanyaan minggu ini: bagaimana jika ingin penalti yang *lebih sensitif* pada ekor ekstrem (heavy-tail, catastrophic losses)?

EVaR (Ahmadi-Javid, 2012): ukuran risiko koheren berbasis momen eksponensial (Chernoff bound) yang cenderung lebih konservatif daripada CVaR.

Aplikasi: pricing reasuransi, capital requirement (RBC/Solvency), operasi dengan rare failures, dan kebijakan RL yang lebih robust.

# Konsep Dasar EVaR

**Definisi (Ahmadi-Javid, 2012).** Untuk kerugian acak  $X$  dan level  $\alpha \in (0, 1)$ :

$$\text{EVaR}_\alpha(X) = \inf_{\lambda > 0} \frac{1}{\lambda} \left( \log \mathbb{E}[e^{\lambda X}] - \log(1 - \alpha) \right).$$

- Koheren (subadditive, monotone, translation invariant, positive homogeneous).
- Bound eksponensial atas kuantil (Chernoff):  $\text{VaR}_\alpha(X) \leq \text{EVaR}_\alpha(X)$ .
- Lebih konservatif pada ekor berat dibanding CVaR, terutama saat *mgf/cumulant* mendominasi.

## Perbandingan Singkat: VaR, CVaR, EVaR

- **VaR**: kuantil; tidak selalu koheren; tidak melihat keparahan di atas ambang.
- **CVaR**: ekspektasi bersyarat di atas VaR; koheren; sensitif ekor tapi linear.
- **EVaR**: berbasis mgf; koheren; penalti eksponensial pada ekor; sering memberikan nilai lebih tinggi (konservatif).

**Intuisi visual:** CVaR = rata-rata area ekor; EVaR = tilting eksponensial yang menekankan skenario sangat buruk.

# EVaR dalam Reinforcement Learning

**Risk-neutral RL:**  $\max_{\pi_{\theta}} J(\theta) = \mathbb{E}_{\pi_{\theta}}[R]$ .

**EVaR-RL:**  $\max_{\pi_{\theta}} J_{\alpha}^{\text{EVaR}}(\theta) = -\text{EVaR}_{\alpha}(-R)$  (mengambil kerugian  $X = -R$ ).

**Bentuk optimisasi:**

$$J_{\alpha}^{\text{EVaR}}(\theta) = \sup_{\lambda > 0} \left\{ -\frac{1}{\lambda} \left( \log \mathbb{E}_{\pi_{\theta}}[e^{-\lambda R}] - \log(1 - \alpha) \right) \right\}.$$

Interpretasi: memilih kebijakan yang memaksimalkan performa *worst-case tilted* oleh eksponential tilting terhadap return.

# Policy Gradient untuk EVaR: Surrogate Dua-Variabel

Definisikan

$$\mathcal{L}_\alpha(\theta, \lambda) = -\frac{1}{\lambda} \left( \log \mathbb{E}_{\pi_\theta} [e^{-\lambda R}] - \log(1 - \alpha) \right), \quad \lambda > 0.$$

Maka  $J_\alpha^{\text{EVaR}}(\theta) = \sup_{\lambda > 0} \mathcal{L}_\alpha(\theta, \lambda)$ .

$\mathcal{L}_\alpha(\theta, \lambda)$  adalah fungsi surrogate dua variabel, yaitu fungsi antara parameter kebijakan  $\theta$  dan parameter eksponensial  $\lambda$ . Fungsi ini merupakan fungsi objektif yang mengaproksimasi dan merepresentasikan EVaR.

- Untuk setiap nilai  $\lambda$ , fungsi ini memberikan ukuran risiko berbasis EVaR dari kebijakan  $\pi_\theta$ .
- Proses optimisasi dua langkah dilakukan :
  - 1 Naikkan  $\theta$  agar kebijakan memaksimalkan return "ter aman" menurut EVaR.
  - 2 Naikkan  $\lambda$  agar menemukan penalti eksponensial optimal untuk tail risk.
- disebut surrogate objective karena digunakan untuk menghitung gradien (policy gradient) secara praktis.

## Gradien (REINFORCE-Style):

$$\nabla_{\theta} \mathcal{L}_{\alpha}(\theta, \lambda) = \frac{\mathbb{E}_{\pi_{\theta}} [e^{-\lambda R} \nabla_{\theta} \log \pi_{\theta}(a|s)]}{\mathbb{E}_{\pi_{\theta}} [e^{-\lambda R}]}.$$

$$\partial_{\lambda} \mathcal{L}_{\alpha}(\theta, \lambda) = \frac{1}{\lambda^2} \left( \log \mathbb{E}[e^{-\lambda R}] - \log(1 - \alpha) \right) + \frac{\mathbb{E}[R e^{-\lambda R}]}{\lambda \mathbb{E}[e^{-\lambda R}]}.$$

(diturunkan via turunan log-mgf). Optimisasi *alternating*: naikkan  $\theta$  (policy gradient), naikkan  $\lambda$  (penalti ekor).

- Untuk  $\nabla_{\theta} \mathcal{L}_{\alpha}(\theta, \lambda)$  pada RL biasa, bobot tiap episode  $\approx R$ . Sedangkan di sini bobotnya  $\approx e^{\lambda R}$ , jadi semakin kecil bobot penalti maka policy menjadi lebih risk-averse.
- Untuk  $\partial_{\lambda} \mathcal{L}_{\alpha}(\theta, \lambda)$  merupakan turunan yang mengatur sensitivitas penalti eksponensial terhadap tail risk. Jika  $\lambda$  meningkat maka penalti terhadap risiko ekstrem makin kuat.
- Jadi,
  - Naikkan  $\theta$  = perbarui kebijakan (mencari policy terbaik untuk risiko saat ini).
  - Naikkan  $\lambda$  = perbarui tingkat penalti (mencari penalti optimal pada ekor distribusi).



# Stabilisasi Praktis

Meskipun EVaR-PG secara teoritis kuat, implementasinya tetap harus berhati-hati supaya :

- Gardien tidak meledak dengan menggunakan baseline/advantage untuk reduksi varians: ganti  $R$  dengan  $A = R - b$ .
- Eksponen tidak overflow dengan melakukan clipping  $\lambda$  dan kontrol numerik untuk eksponen (log-sum-exp trick).
- Sampling efisien melalui mini-batch Monte Carlo menggunakan reweighting : ; gunakan reweighting  $w_i = \exp(-\lambda R_i)$ .
- Hubungan ke CVaR-PG: saat penalti eksponensial kecil, perilaku mendekati fokus lower-tail.

# Demo dan Implementasi

## RL Mini-Project: CliffWalking (tail risk)

- Bandingkan **REINFORCE** (risk-neutral) vs **EVaR-PG** pada  $\alpha \in \{0.95, 0.99\}$ .
- **Metrik**: mean return,  $\text{VaR}_\alpha$ ,  $\text{CVaR}_\alpha$ , *frekuensi jatuh (cliff)*.
- Ekspektasi: EVaR-PG memilih rute lebih aman; mean return bisa sedikit turun namun tail risk berkurang tajam.

## Studi Data Pricing (loss domain)

- *Light-tail* (Lognormal) vs *Heavy-tail* (Lognormal/Pareto mix).
- Hitung premi berbasis Expected Loss,  $\text{VaR}_\alpha$ ,  $\text{CVaR}_\alpha$ , dan **EVaR** untuk  $\alpha \in \{0.95, 0.99\}$ .
- Visual: histogram (loss) + garis  $\text{VaR}/\text{CVaR}$ ; tabel perbandingan premi.

## Catatan domain :

- *Pricing* pakai loss  $\rightarrow$  tail yang dipakai upper tail.
- *RL* pakai retur  $\rightarrow$  tail yang dipakai lower tail.

# Panduan Diskusi: Interpretasi Hasil Praktikum

## Bagian A Pricing (Loss Domain)

- Bandingkan nilai **Expected**, **VaR**, **CVaR**, dan **EVaR** pada *light-tail* dan *heavy-tail*.
  - Apakah urutannya konsisten: Expected ; VaR ; CVaR ; EVaR?
  - Mengapa perbedaan antar ukuran risiko lebih besar pada *heavy-tail*?
- Diskusikan implikasi praktis: bagaimana EVaR memengaruhi besarnya **premi asuransi** dan **cadangan modal**?
- Apa konsekuensi jika perusahaan asuransi hanya memakai Expected atau CVaR untuk risiko katastrofik?

## A. Perbaikan Environment

- Saat menginjak *cliff*: beri  $r = -100$ , **kembalikan ke start, lanjutkan episode** (`done = False`).
- Simpan `self.s = ns` (hapus logika reset saat done).
- Disarankan: `max_steps = 200`,  $\gamma = 0.99$ ; opsional batasi jatuh/episode ( $\leq 5$ ).

## B. Stabilisasi Training

- **Baseline** per episode:  $b = \text{mean}(G)$ .
- **REINFORCE (risk-neutral)**: bobot gradien  $w = (G - b)$ .
- **EVaR-PG ( $\lambda$  tetap, versi stabil)**:  $w = -e^{-\lambda G} + e^{-\lambda b}$ .
- Setelan awal:  $1r \in [0.005, 0.01]$ ,  $\lambda \in [0.005, 0.01]$ ; opsional tambah *entropy bonus* kecil.

## C. Output Minimum

- Plot *episode returns* (2 kurva) dan laporkan `fall_freq`.
- Tabel ringkas: mean return, VaR/CVaR @  $\{0.95, 0.99\}$ .

# Panduan Diskusi: Interpretasi Hasil Praktikum (lanj.)

## Bagian B RL Mini-Project: CliffWalking (Return Domain)

- Bandingkan hasil **REINFORCE (risk-neutral)** dan **EVaR-PG**.
  - Bagaimana perbedaan **mean return**,  **$\text{VaR}_\alpha$** ,  **$\text{CVaR}_\alpha$** , dan **frekuensi jatuh**?
  - Apakah EVaR-PG benar-benar menghasilkan *policy* yang lebih aman?
- Diskusikan pengaruh nilai  $\alpha$ :
  - Mengapa saat  $\alpha$  meningkat dari 0.95 ke 0.99, strategi EVaR-PG menjadi lebih konservatif?
  - Apa hubungan  $\alpha$  dengan tingkat kepercayaan atau toleransi risiko?
- Hubungkan kembali ke teori:
  - Bagaimana peran parameter  $\lambda$  dalam membentuk tingkat keengganan terhadap risiko ekstrem?
  - Apakah perilaku EVaR-PG mendekati CVaR-PG saat  $\lambda$  kecil?

# Panduan Diskusi: Sintesis dan Refleksi

- Apa kesamaan konsep antara **EVaR dalam Pricing** dan **EVaR dalam Reinforcement Learning**?
- Bagaimana ukuran risiko seperti EVaR membantu pengambilan keputusan dalam:
  - penentuan premi dan cadangan di asuransi, dan
  - pembentukan kebijakan (policy) dalam lingkungan berisiko?
- Jika Anda menjadi **aktuaris atau perancang kebijakan**, kapan memilih EVaR dibanding CVaR?
- Bagaimana konsep ini bisa diperluas untuk model lain misalnya reinsurance, portfolio optimization, atau risk-sensitive control?

# Analisis Trade-off

- Risk-neutral: expected gain tinggi, tail risk tinggi.
- CVaR: menurunkan tail risk; masih linear terhadap ekor.
- EVaR: penalti eksponensial pada ekor  $\Rightarrow$  strategi paling konservatif.
- Sensitivitas terhadap  $\alpha$  dan  $\lambda$ : semakin mendekati 1 (atau besar tilt), semakin konservatif dan stabil.
- Implikasi industri: premi lebih tinggi, modal lebih besar, probabilitas ruin lebih kecil.

# Referensi Singkat

Ahmadi-Javid, A. (2012). Entropic Value-at-Risk: A New Coherent Risk Measure. JOTA.

Tamar, A., et al. (2015). Policy Gradients with Variance Related Risk Criteria.

Chow, Y., et al. (2015–2017). Risk-Sensitive/CVaR MDPs.

Frees, E. W. (dataset) Health Insurance.