Verification of Discrete Upper Confidence Bound algorithm in Isabelle/HOL

Arjan Faber

February 6, 2026

Abstract

This project formally verifies the Upper Confidence Bound (UCB) algorithm in Isabelle/Higher-order Logic (HOL), focusing on its probabilistic guarantees and regret bounds. The work extends Isabelle/HOLs probabilistic framework and explores verification of discrete-time bandit models following [1]. This research advances the formal verification of probabilistic algorithms in reinforcement learning.

theory *MSc-Project-Discrete-Prop15-1*
imports
HOL−Probability.Probability

begin

locale *bandit-problem* =
fixes $A :: 'a\ set$
and $\mu :: 'a \Rightarrow real$
and $a\text{-}star :: 'a$
assumes *finite-arms*: $finite\ A$
and *a-star-in-A*: $a\text{-}star \in A$
and *optimal-arm*: $\forall\, a \in A.\ \mu\ a\text{-}star \geq \mu\ a$
begin

definition $\Delta :: 'a \Rightarrow real$ **where**
$\Delta\ a = \mu\ a\text{-}star - \mu\ a$
end

locale *bandit-policy* = *bandit-problem* + *prob-space* +
fixes $\Omega :: 'b\ set$
and $\mathcal{F} :: 'b\ set\ set$
and $\pi :: nat \Rightarrow 'b \Rightarrow 'a$
and $N\text{-}n :: nat \Rightarrow 'a \Rightarrow 'b \Rightarrow nat$
assumes *measurable-policy*: $\forall\, t.\ \pi\ t \in measurable\ M\ (count\text{-}space\ A)$
and *N-n-def*: $\forall\, n\ a\ \omega.\ N\text{-}n\ n\ a\ \omega = card\ \{t \in \{0..<n\}.\ \pi\ (t+1)\ \omega = a\}$
and *count-assm-pointwise*: $\forall\, n\ \omega.\ (\sum a \in A.\ real\ (N\text{-}n\ n\ a\ \omega)) = real\ n$

begin

definition *R-n* :: *nat* ⇒ *'b* ⇒ *real* **where**
R-n n ω = *n* ∗ *μ a-star* − (∑ *a* ∈ *A*. *μ a* ∗ *real* (*N-n n a ω*))

lemma *regret-decomposition-pointwise*:
fixes *n* :: *nat* **and** *ω* :: *'b*
assumes *n-count-assm-pointwise*: (∑ *a*∈*A*. *real* (*N-n n a ω*)) = *real n*
shows *R-n n ω* = (∑ *a* ∈ *A*. Δ *a* ∗ *real* (*N-n n a ω*))
⟨*proof*⟩

lemma *integrable-const-fun*:
assumes *finite-measure M*
shows *integrable M* (λ*x*. *c*)
⟨*proof*⟩

lemma *expected-regret*:
assumes *finite A*
and ∀ *a* ∈ *A*. *integrable M* (λ*ω*. *real* (*N-n n a ω*))
shows *expectation* (λ*ω*. *R-n n ω*) = (∑ *a*∈*A*. Δ *a* ∗ *expectation* (λ*ω*. *real* (*N-n n a ω*)))
⟨*proof*⟩

end

end
theory *Discrete-UCB-Step1*
imports *MSc-Project-Discrete-Prop15-1*

begin

locale *bandit-policy* = *bandit-problem* + *prob-space* +
fixes Ω :: *'b set*
and ℱ :: *'b set set*
and *ω* :: *'b*
and *π* :: *nat* ⇒ *'b* ⇒ *'a*
and *N-n* :: *nat* ⇒ *'a* ⇒ *'b* ⇒ *nat*
assumes *measurable-policy*: ∀ *t*. *π t* ∈ *measurable M* (*count-space A*)
and *N-n-def*: ∀ *n a ω*. *N-n n a ω* = *card* {*t* ∈ {*0*..<*n*}. *π* (*t*+*1*) *ω* = *a*}
and *count-assm-pointwise*: ∀ *n ω*. (∑ *a* ∈ *A*. *real* (*N-n n a ω*)) = *real n*
begin

lemma *union-eq*:
fixes *a* :: *'a* **and** *n k* :: *nat*
assumes *k* ≤ *n*
shows {*t*. *t* < *n* ∧ *π* (*t*+*1*) *ω* = *a*} = {*t*. *t* < *k* ∧ *π* (*t*+*1*) *ω* = *a*} ∪ {*t*. *k* ≤ *t*

$\wedge\ t < n \wedge \pi\ (t{+}1)\ \omega = a\}$
⟨*proof*⟩

lemma *cardinality-indic-eq*:
fixes $I :: nat \Rightarrow bool$
assumes *finite* $\{t.\ k \leq t \wedge t < n\}$
shows *card* $\{t.\ k \leq t \wedge t < n \wedge \pi\ (t{+}1)\ \omega = a \wedge I\ t\} = (\sum\ t = k..{<}n.\ if\ \pi\ (t{+}1)\ \omega = a \wedge I\ t\ then\ 1\ else\ 0)$
⟨*proof*⟩

lemma *ge-rewrite*: $(x{::}real) \geq y \Longrightarrow y \leq x$ ⟨*proof*⟩

lemma *Nn-expression*:
fixes $a :: 'a$ **and** $s :: nat \Rightarrow real$
and $k :: nat$ **and** $n :: nat$
assumes $a \in A$
and $k \leq n$
and $0 < n$
and $\forall\ t \in \{0..n\}.\ 0 < s\ t$
and $\forall\, t < n - 1.\ \ s\ t \leq s\ (t + 1)$
and *init-play-once*: $\forall\, \omega.\ a \in A \longrightarrow N{-}n\ k\ a\ \omega = 1$
and *finite-played-sets*:
finite $\{t.\ t < n \wedge \pi\ (t{+}1)\ \omega = a\}$
finite $\{t.\ t < k \wedge \pi\ (t{+}1)\ \omega = a\}$
finite $\{t.\ k \leq t \wedge t < n \wedge \pi\ (t{+}1)\ \omega = a\}$
shows
$(N{-}n\ n\ a\ \omega) = 1 + (\sum\ t = k..{<}n.\ if\ \pi\ (t{+}1)\ \omega = a \wedge real\ (N{-}n\ t\ a\ \omega) < s\ t\ then\ 1\ else\ 0) +$
$(\sum\ t = k..{<}n.\ if\ \pi\ (t{+}1)\ \omega = a \wedge real\ (N{-}n\ t\ a\ \omega) \geq s\ t\ then\ 1\ else\ 0)$
⟨*proof*⟩

lemma *upper-bound-expression-contradiction*:
fixes $a :: 'a$ **and** $s :: nat \Rightarrow real$
and $k :: nat$ **and** $n :: nat$
and *s-n-nat* :: *nat*
assumes $a \in A$
and $k \leq n$
and $0 < n$
and *non-neg-s*: $\forall\ t \in \{0..n\}.\ 0 < s\ t$
and *base-le*: $s\ 0 \leq s\ 1$
and *non-dec*: $\forall\, t < n - 1.\ s\ t \leq s\ (t + 1)$
and *s-mono*: $\bigwedge t.\ k \leq t \wedge t \leq n \Longrightarrow s\ t \leq s\ n$
and *init-play-once*: $\forall\, \omega.\ a \in A \longrightarrow N{-}n\ k\ a\ \omega = 1$
and *finite-played-sets*:
finite $\{t.\ t < n \wedge \pi\ (t{+}1)\ \omega = a\}$
finite $\{t.\ t < k \wedge \pi\ (t{+}1)\ \omega = a\}$
finite $\{t.\ k \leq t \wedge t < n \wedge \pi\ (t{+}1)\ \omega = a\}$
and *xs-sorted-def*: $xs = sorted\text{-}list\text{-}of\text{-}set\ \{t \in \{k..{<}n\}.\ \pi\ (t{+}1)\ \omega = a \wedge real$

$(N\text{-}n\ t\ a\ \omega) < s\ t\}$
and *s-nat-def*: $s\text{-}n\text{-}nat = nat\ (\lfloor s\ n \rfloor)$
and *len-bound-def*: $s\text{-}n\text{-}nat < length\ xs$
and *distinct-xs*: $distinct\ xs$
and *gt-ineq*: $length\ xs + 1 > \lfloor s\ n \rfloor$
and *N-n-increasing-with-plays*:
$\forall\ t\ t'.\ k \leq t \wedge t < t' \wedge \pi\ (t{+}1)\ \omega = a \wedge \pi\ (t'{+}1)\ \omega = a \longrightarrow N\text{-}n\ t'\ a\ \omega \geq N\text{-}n\ t\ a\ \omega + 1$
and *neg*: $1 + real\ (\sum\ t{=}k..{<}n.\ if\ \pi\ (t{+}1)\ \omega = a \wedge real\ (N\text{-}n\ t\ a\ \omega) < s\ t\ then\ 1\ else\ 0) > s\ n$
and *t-hat* $\in$ *set xs*
and $t\text{-}hat = xs\ !\ s\text{-}n\text{-}nat$
and $real\ (N\text{-}n\ t\text{-}hat\ a\ \omega) \geq real\ s\text{-}n\text{-}nat + 1$

shows $(real\ (N\text{-}n\ t\text{-}hat\ a\ \omega) \geq \lfloor s\ n \rfloor + 1) \wedge (\pi\ (t\text{-}hat{+}1)\ \omega = a \wedge (real\ (N\text{-}n\ t\text{-}hat\ a\ \omega) < s\ t\text{-}hat))$

⟨*proof*⟩

lemma *Nn-upper-bound*:
fixes $a :: 'a$ **and** $s :: nat \Rightarrow real$
and $k :: nat$ **and** $n :: nat$
assumes *asm*: $real(1 + (\sum t = k..{<}n.\ if\ \pi\ (t{+}1)\ \omega = a \wedge real\ (N\text{-}n\ t\ a\ \omega) < s\ t\ then\ 1\ else\ 0)) \leq s\ n$
and *a-in-A*: $a \in A$
and *k-le-n*: $k \leq n$
and *n-pos*: $0 < n$
and *s-pos*: $\forall\ t \in \{0..n\}.\ 0 < s\ t$
and *s-nondec*: $\forall\ t < n - 1.\ s\ t \leq s\ (t + 1)$
and *init-play-once*: $\forall \omega.\ a \in A \longrightarrow N\text{-}n\ k\ a\ \omega = 1$
and *finite-played-sets-1*: $finite\ \{t.\ t < n \wedge \pi\ (t{+}1)\ \omega = a\}$
and *finite-played-sets-2*: $finite\ \{t.\ t < k \wedge \pi\ (t{+}1)\ \omega = a\}$
and *finite-played-sets-3*: $finite\ \{t.\ k \leq t \wedge t < n \wedge \pi\ (t{+}1)\ \omega = a\}$
shows $real(N\text{-}n\ n\ a\ \omega) \leq s\ n + real((\sum t = k..{<}n.\ if\ \pi\ (t{+}1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0))$
⟨*proof*⟩

theorem *ENn-upper-bound*:
assumes
a-in-A: $a \in A$
and *k-le-n*: $k \leq n$
and *n-pos*: $0 < n$

and *s-pos*: $\forall\, t \in \{0..n\}.\ 0 < s\ t$
and *s-nondec*: $\forall\, t < n.\ s\ t \leq s\ (t + 1)$
and *init-play-once*: $\forall\, \omega.\ a \in A \longrightarrow N\text{-}n\ k\ a\ \omega = 1$
and *integrable-Nn*: *integrable* M $(\lambda\omega.\ real\ (N\text{-}n\ n\ a\ \omega))$
and *integrable-rhs-sum*: *integrable* M $(\lambda\omega.\ s\ n + (\sum t = k..<n.\ if\ \pi\ (t+1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0))$
and *integrable-s*: *integrable* M $(\lambda\omega.\ s\ n)$
and *integrable-indicator-sum*: *integrable* M $(\lambda\omega.\ \sum t = k..<n.\ if\ \pi\ (t+1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0)$
and *linearity*: $integral^L\ M\ (\lambda\omega.\ s\ n + (\sum t = k..<n.\ if\ \pi\ (t+1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0)) =$
$integral^L\ M\ (\lambda\omega.\ s\ n) + integral^L\ M\ (\lambda\omega.\ \sum t = k..<n.\ if\ \pi\ (t+1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0)$
and *pointwise-bound*: $real\ (N\text{-}n\ n\ a\ \omega) \leq s\ n + (\sum t = k..<n.\ if\ \pi\ (t+1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0)$
and *mono-intgrl*: $integral^L\ M\ (\lambda\omega.\ real\ (N\text{-}n\ n\ a\ \omega)) \leq integral^L\ M\ (\lambda\omega.\ s\ n + (\sum t = k..<n.\ if\ \pi\ (t+1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0))$
shows
$expectation\ (\lambda\omega.\ real\ (N\text{-}n\ n\ a\ \omega)) \leq$
$s\ n + expectation\ (\lambda\omega.\ (\sum t = k..<n.\ \ if\ \pi\ (t+1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0))$
⟨*proof*⟩

end
end
theory *Discrete-UCB-Step2*
imports *Discrete-UCB-Step1*

begin

locale *bandit-policy* = *bandit-problem* + *prob-space* +
fixes Ω :: *'b set*
and $\mathcal{F}$ :: *'b set set*
and ω :: *'b*
fixes π :: *nat* ⇒ *'b* ⇒ *'a*
and *N-n* :: *nat* ⇒ *'a* ⇒ *'b* ⇒ *nat*
and *Z* :: *nat* ⇒ *'a* ⇒ *'b* ⇒ *real*
and δ :: *real*
and *q* :: *real*
assumes *finite-A*: *finite A*
and *a-in-A*: $a \in A$
and *measurable-policy*: $\forall\, t.\ \pi\ t \in measurable\ M\ (count\text{-}space\ A)$
and *N-n-def*: $\forall\, n\ a\ b.\ N\text{-}n\ n\ a\ b = card\ \{t \in \{0..<n\}.\ \pi\ (t+1)\ b = a\}$
and *δ-pos*: $0 < \delta$
and *δ-less1*: $\delta < 1$
and *q-pos*: $q > 0$

begin

definition *sample-mean-Z* :: *nat* ⇒ *'a* ⇒ *'b* ⇒ *real* **where**
sample-mean-Z n a b ≡ (*1 / real n*) ∗ (∑ *i*<*n*. *Z i a b*)

definition *M-val* :: *nat* ⇒ *'a* ⇒ *'b* ⇒ *real* **where**
M-val t a b ≡ (*if N-n* (*t*+*1*) *a b* = *0 then 0*
else (∑ *s* < *t*. *if* π *s b* = *a then Z s a b else 0*) / *real* (*N-n t a b*))

definition *U* :: *nat* ⇒ *'a* ⇒ *'b* ⇒ *real* **where**
U t a b ≡ *M-val t a b* + *q* ∗ *sqrt* (*ln* (*1 / δ*) / (*2* ∗ *real* (*max 1* (*N-n t a b*))))

definition *A-t-plus-1* :: *nat* ⇒ *'b* ⇒ *'a* **where**
A-t-plus-1 t b ≡ (*SOME a*. *a* ∈ *A* ∧ (∀ *a'*. *a'* ∈ *A* ⟶ *U t a b* ≥ *U t a' b*))

lemma (**in** *finite-measure*) *finite-measure-mono*:
assumes *A* ⊆ *B B* ∈*sets M* **shows** *measure M A* ≤ *measure M B*
⟨*proof*⟩

theorem *union-bound*:
fixes *E F G* :: *'b set*
assumes *E* ⊆ *F* ∪ *G*
and *E* ∈ *events F* ∈ *events G* ∈ *events*
shows *prob E* ≤ *prob F* + *prob G*
⟨*proof*⟩

theorem *hoeffding-iid-bound-ge-general*:
fixes *a* :: *'a* **and** *n* :: *nat* **and** ε :: *real* **and** *μ-hat* :: *real* **and** *l u* :: *real*
assumes *a-in*: *a* ∈ *A*
and *eps-pos*: ε ≥ *0*
and *bounds*: ∀ *i* < *n*. ∀ ω ∈ Ω. *l* ≤ *Z i a* ω ∧ *Z i a* ω ≤ *u*
and *mu-def*: *μ-hat* = (∑ *i* < *n*. *expectation* (λω. *Z i a* ω))
and *u* − *l* ≠ *0*
and *n-pos*: *n* > *0*
and *space-M*: *space M* = Ω
and *sets-M*: *sets M* = *F*
and *indep*: *indep-vars* (λ-. *borel*) (λ*i*. (λω. *Z i a* ω)) {*i*. *i* < *n*}
and *rv*: ∀ *i*<*n*. *random-variable borel* (λω. *Z i a* ω)

shows *prob* {ω ∈ Ω. (∑ *i* < *n*. *Z i a* ω) ≥ *μ-hat* + ε}
≤ *exp* (− *2* ∗ ε^*2* / (*real n* ∗ (*u* − *l*)^*2*))
⟨*proof*⟩

theorem *hoeffding-iid-bound-le-general*:
fixes *a* :: *'a* **and** *n* :: *nat* **and** ε :: *real* **and** *μ-hat* :: *real* **and** *l u* :: *real*
assumes *a-in*: *a* ∈ *A*
and *eps-pos*: ε ≥ *0*

and *bounds*: $\forall\, i < n.\ \forall \omega \in \Omega.\ l \leq Z\ i\ a\ \omega \wedge Z\ i\ a\ \omega \leq u$
and *mu-def*: $\mu\text{-}hat = (\sum\ i < n.\ expectation\ (\lambda\omega.\ Z\ i\ a\ \omega))$
and $u - l \neq 0$
and *n-pos*: $n > 0$
and *space-M*: $space\ M = \Omega$
and *sets-M*: $sets\ M = \mathcal{F}$
and *indep*: $indep\text{-}vars\ (\lambda\text{-}.\ borel)\ (\lambda i.\ (\lambda\omega.\ Z\ i\ a\ \omega))\ \{i.\ i < n\}$
and *rv*: $\forall\, i{<}n.\ random\text{-}variable\ borel\ (\lambda\omega.\ Z\ i\ a\ \omega)$
shows $prob\ \{\omega \in \Omega.\ (\sum\ i < n.\ Z\ i\ a\ \omega) \leq \mu\text{-}hat - \varepsilon\}$
$\leq exp\ (-\ 2 * \varepsilon\hat{}2\ /\ (real\ n * (u - l)\hat{}2))$
⟨*proof*⟩

theorem *hoeffding-iid-ge-delta-bound*:
fixes $a :: 'a$ **and** $n :: nat$ **and** $\delta\text{-}hat :: real$ **and** $\mu\text{-}hat :: real$ **and** $l\ u :: real$
assumes *a-in*: $a \in A$
and *delta-bound*: $0 < \delta\text{-}hat\ \delta\text{-}hat \leq 1$
and *bounds*: $\forall\, i{<}n.\ \forall \omega{\in}\Omega.\ l \leq Z\ i\ a\ \omega \wedge Z\ i\ a\ \omega \leq u$
and *mu-def*: $\mu\text{-}hat = (\sum i{<}n.\ expectation\ (\lambda\omega.\ Z\ i\ a\ \omega))$
and *n-pos*: $n > 0$
and *eps-pos*: $\varepsilon \geq 0$
and *u-minus-l-nonzero*: $u - l \neq 0$
and *space-M*: $space\ M = \Omega$
and *sets-M*: $sets\ M = \mathcal{F}$
and *indep*: $indep\text{-}vars\ (\lambda\text{-}.\ borel)\ (\lambda i.\ (\lambda\omega.\ Z\ i\ a\ \omega))\ \{i.\ i < n\}$
and *rv*: $\forall\, i{<}n.\ random\text{-}variable\ borel\ (\lambda\omega.\ Z\ i\ a\ \omega)$
and *eps-expression*: $\varepsilon = sqrt\ ((real\ n * (u - l)\hat{}2 * ln\ (1\ /\ \delta\text{-}hat))\ /\ 2)$
shows $prob\ \{\omega \in \Omega.\ (\sum i{<}n.\ Z\ i\ a\ \omega) \geq \mu\text{-}hat + \varepsilon\} \leq \delta\text{-}hat \wedge$
$prob\ \{\omega \in \Omega.\ (\sum i{<}n.\ Z\ i\ a\ \omega) \leq \mu\text{-}hat - \varepsilon\} \leq \delta\text{-}hat$
⟨*proof*⟩

lemma *add-le-iff*:
fixes $x\ y\ z :: real$
shows $x \leq y - z \longleftrightarrow x - y \leq -z$
⟨*proof*⟩
lemma *max-Suc-0-eq-1*: $max\ (Suc\ 0)\ x = max\ 1\ x$
⟨*proof*⟩

theorem *ucb-suboptimal-bound-set*:
fixes $t :: nat$
and $a :: 'a$
and $\Delta :: 'a \Rightarrow real$
assumes *finite-A*: $finite\ A$
and *a-in-A*: $a \in A$
and *a-star-in-A*: $a\text{-}star \in A$
and *argmax-exists*: $A \neq \{\}$
and *subopt-gap*: $\Delta\ a > 0$
and *a-not-opt*: $\exists\, a'.\ a' \in A \wedge \Delta\ a > 0$
and *delta-a*: $\Delta\ a = \mu\ a\text{-}star - \mu\ a$

and *ω-in-Ω*: $\omega \in \Omega$
and *asm*: $\omega \in \{\omega \in \Omega.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a\}$
and *setopt*: $\forall \omega \in \Omega.\ \exists\, a\text{-}max \in A.\ \forall\, b \in A.\ U\ t\ b\ \omega \leq U\ t\ a\text{-}max\ \omega$
and *A-t-plus-1-maximizes*:
$\bigwedge t\ \omega\ a.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a \Longrightarrow a \in A \wedge (\forall\, b \in A.\ U\ t\ a\ \omega \geq U\ t\ b\ \omega)$
shows $\{\omega \in \Omega.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a\} \subseteq$
$\{\omega \in \Omega.\ U\ t\ a\text{-}star\ \omega \leq \mu\ a\text{-}star\} \cup \{\omega \in \Omega.\ \mu\ a\text{-}star \leq U\ t\ a\ \omega\}$
⟨*proof*⟩

theorem *ucb-suboptimal-bound-prob-statement*:
fixes t :: *nat* **and** a :: $'a$ **and** Δ :: $'a \Rightarrow real$
assumes *finite-A*: *finite A*
and *a-star-in-A*: $a\text{-}star \in A$
and *argmax-exists*: $A \neq \{\}$
and *subopt-gap*: $\Delta\ a > 0$
and *a-not-opt*: $\exists\, a'.\ a' \in A \wedge \Delta\ a > 0$
and *ω-in-Ω*: $\omega \in \Omega$
and *asm*: $\omega \in \{\omega \in \Omega.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a\}$
and *setopt*: $\forall \omega \in \Omega.\ \exists\, a\text{-}max \in A.\ \forall\, b \in A.\ U\ t\ b\ \omega \leq U\ t\ a\text{-}max\ \omega$
and *A-t-plus-1-maximizes*:
$\bigwedge t\ \omega\ a.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a \Longrightarrow a \in A \wedge (\forall\, b \in A.\ U\ t\ a\ \omega \geq U\ t\ b\ \omega)$
and *a-in-A*: $a \in A$
and *omega-in*: $\omega \in \Omega$
and *subopt-gap*: $\Delta\ a > 0$
and *delta-a*: $\Delta\ a = \mu\ a\text{-}star - \mu\ a$
and *H-def*: $H = (2 * ln\ (1\ /\ \delta))\ /\ (\Delta\ a)\hat{}2$
and *E-def*: $E = \{\omega \in \Omega.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a\} \cap \{\omega \in \Omega.\ H \leq real\ (N\text{-}n\ t\ a\ \omega)\}$
and *F-def*: $F = \{\omega \in \Omega.\ U\ t\ a\text{-}star\ \omega \leq \mu\ a\text{-}star\} \cap \{\omega \in \Omega.\ H \leq real\ (N\text{-}n\ t\ a\ \omega)\}$
and *G-def*: $G = \{\omega \in \Omega.\ \mu\ a\text{-}star \leq U\ t\ a\ \omega\} \cap \{\omega \in \Omega.\ H \leq real\ (N\text{-}n\ t\ a\ \omega)\}$
and *meas-sets*: $E \in sets\ M\ F \in sets\ M\ G \in sets\ M$
and *prob-inter*: $prob\ \ (F \cap G) \equiv enn2real\ (emeasure\ M\ (F \cap G))$

shows $prob\ \ (\{\omega \in \Omega.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a\} \cap \{\omega \in \Omega.\ H \leq real\ (N\text{-}n\ t\ a\ \omega)\}) \leq$
$prob\ \ (\{\omega \in \Omega.\ U\ t\ a\text{-}star\ \omega \leq \mu\ a\text{-}star\} \cap \{\omega \in \Omega.\ H \leq real\ (N\text{-}n\ t\ a\ \omega)\})$
$+$
$prob\ \ (\{\omega \in \Omega.\ U\ t\ a\ \omega \geq \mu\ a\text{-}star\} \cap \{\omega \in \Omega.\ H \leq real\ (N\text{-}n\ t\ a\ \omega)\})$
⟨*proof*⟩

lemma *U-le-μ-pointwise*:
$U\ t\ a\text{-}star\ \omega \leq \mu\ a\text{-}star \longleftrightarrow$
$M\text{-}val\ t\ a\text{-}star\ \omega - \mu\ a\text{-}star \leq$
$-\ q * sqrt\ (ln\ (1\ /\ \delta)\ /\ (2 * real\ (max\ 1\ (N\text{-}n\ t\ a\text{-}star\ \omega))))$
⟨*proof*⟩

lemma *U-ge-μ-pointwise*:
assumes *delta-a*: $\Delta\ a = \mu\ a\text{-}star - \mu\ a$
shows
$U\ t\ a\ \omega \geq \mu\ a\text{-}star \longleftrightarrow$
$M\text{-}val\ t\ a\ \omega - \mu\ a \geq \Delta\ a - q * sqrt\ (ln\ (1\ /\ \delta)\ /\ (2 * real\ (max\ 1\ (N\text{-}n\ t\ a\ \omega))))$
⟨*proof*⟩

theorem *hoeffding-iid--bound-le*:
fixes $a :: 'a$ **and** $n :: nat$ **and** $\varepsilon :: real$ **and** $\mu\text{-}hat :: real$ **and** $l\text{-}hat\ u\text{-}hat :: real$
and $I :: nat\ set$
and $X\text{-}new :: nat \Rightarrow 'b \Rightarrow real$
and $a\text{-}bound\ b\text{-}bound :: nat \Rightarrow real$
assumes *a-in*: $a \in A$
and $b \in \Omega$
and *eps-pos*: $\varepsilon \geq 0$
and *eps*: $\varepsilon = abs\ (u\text{-}hat - l\text{-}hat) * sqrt\ (((real\ n)\ /\ 2) * ln\ (1\ /\ \delta))$
and $\delta \geq 0 \wedge \delta \leq 1$
and *t-eq-n*: $t = n$
and $c > 0$
and *bounds*: $\forall j < t.\ \forall \omega \in \Omega.\ \forall\ a \in A.\ l\text{-}hat \leq Z\text{-}hat\ j\ a\ \omega \wedge Z\text{-}hat\ j\ a\ \omega \leq u\text{-}hat$
and *mu-def*: $\mu\text{-}hat = (\sum\ j < t.\ expectation\ (\lambda\omega.\ Z\text{-}hat\ j\ a\ \omega))$
and $u\text{-}hat - l\text{-}hat \neq 0$
and *t-pos*: $t > 0$
and $\forall t{<}n.\ N\text{-}n\ t\ a\text{-}star\ b > 0$
and *n-pos*: $n > 0$
and $M\text{-}val\ t\ a\text{-}star\ b \equiv (\sum\ s < t.\ if\ \pi\ s\ b = a\text{-}star\ then\ Z\ s\ a\text{-}star\ b\ else\ 0)\ /\ real\ (N\text{-}n\ t\ a\text{-}star\ b)$
and *widths*: $(\sum\ i \in I.\ (b\text{-}bound\ i - a\text{-}bound\ i)\hat{}2) = (real\ n) * (u\text{-}hat - l\text{-}hat)\hat{}2$
and *space-M*: $space\ M = \Omega$
and *sets-M*: $sets\ M = \mathcal{F}$
and *indep*: $indep\text{-}vars\ (\lambda\text{-}.\ borel)\ (\lambda j.\ (\lambda\omega.\ Z\ j\ a\ \omega))\ \{j.\ j < t\}$
and *rv*: $\forall j{<}t.\ random\text{-}variable\ borel\ (\lambda\omega.\ Z\ j\ a\ \omega)$
and $\forall j{<}t.\ Z\text{-}hat\ j\ a\text{-}star\ \omega = c * (if\ \pi\ j\ b = a\text{-}star\ then\ Z\ j\ a\text{-}star\ b\ else\ 0)$
and *indep-interval-bounded-random-variables M I X-new a-bound b-bound*
and *indep-loc*: *indep-interval-bounded-random-variables M* $\{j.\ j < t\}$
$(\lambda j.\ (\lambda\omega.\ Z\text{-}hat\ j\ a\text{-}star\ \omega))$
$(\lambda j.\ l\text{-}hat)\ (\lambda j.\ u\text{-}hat)$
and *H*: *Hoeffding-ineq M* $\{j.\ j < t\}$
$(\lambda j.\ (\lambda\omega.\ Z\text{-}hat\ j\ a\text{-}star\ \omega))$
$(\lambda j.\ l\text{-}hat)\ (\lambda j.\ u\text{-}hat)$
and *sum-integrals-eq*: $(\sum\ j \in \{j.\ j < t\}.\ integral^L\ M\ (\lambda\omega.\ Z\text{-}hat\ j\ a\text{-}star\ \omega)) = \mu\text{-}hat$
and *rewriting*: $prob\ \{\omega \in \Omega.\ (\sum\ j < t.\ Z\text{-}hat\ j\ a\text{-}star\ \omega) - (\sum\ j < t.\ expectation\ (Z\text{-}hat\ j\ a\text{-}star)) \leq -\varepsilon\} =$
$prob\ \{x \in space\ M.\ (\sum\ j < t.\ Z\text{-}hat\ j\ a\text{-}star\ x) \leq (\sum\ j < t.\ expectation\ (Z\text{-}hat\ j\ a\text{-}star)) - \varepsilon\}$

shows *prob* $\{\omega \in \Omega.\ (\sum\ j < n.\ Z\text{-}hat\ j\ a\text{-}star\ \omega) \leq \mu\text{-}hat - \varepsilon\}$
$\leq \delta$
⟨*proof*⟩

theorem *hoeffding-iid--bound-ge*:
fixes $a :: 'a$ **and** $n :: nat$ **and** $\varepsilon :: real$ **and** $\mu\text{-}hat :: real$ **and** *l-hat u-hat* :: *real*
and $I :: nat\ set$
and *X-new* :: $nat \Rightarrow 'b \Rightarrow real$
and *a-bound b-bound* :: $nat \Rightarrow real$
assumes *a-in*: $a \in A$
and $b \in \Omega$
and *eps-pos*: $\varepsilon \geq 0$
and *eps*: $\varepsilon =\ abs\ (u\text{-}hat - l\text{-}hat) * sqrt\ (((real\ n)\ /\ 2) * ln\ (1\ /\ \delta))$
and $\delta \geq 0 \wedge \delta \leq 1$
and *t-eq-n*: $t = n$
and $c > 0$
and *bounds*: $\forall j < t.\ \forall \omega \in \Omega.\ \forall\ a \in A.\ l\text{-}hat \leq Z\text{-}hat\ j\ a\ \omega \wedge Z\text{-}hat\ j\ a\ \omega \leq u\text{-}hat$
and *mu-def*: $\mu\text{-}hat = (\sum\ j < t.\ expectation\ (\lambda\omega.\ Z\text{-}hat\ j\ a\ \omega))$
and $u\text{-}hat - l\text{-}hat \neq 0$
and *t-pos*: $t > 0$
and $\forall\, t{<}n.\ \ N\text{-}n\ t\ a\text{-}star\ b > 0$
and *n-pos*: $n > 0$
and *M-val t a b* $\equiv (\sum\ s < t.\ if\ \pi\ s\ b = a\ then\ Z\ s\ a\ b\ else\ 0)\ /$
$real\ (N\text{-}n\ t\ a\ b)$
and *widths*: $(\sum\ i \in I.\ (b\text{-}bound\ i - a\text{-}bound\ i)\hat{}2) = (real\ n) * (u\text{-}hat - l\text{-}hat)\hat{}2$
and *space-M*: $space\ M = \Omega$
and *sets-M*: $sets\ M\ = \mathcal{F}$
and *indep*: *indep-vars* $(\lambda\text{-}.\ borel)\ (\lambda j.\ (\lambda\omega.\ Z\ j\ a\ \omega))\ \{j.\ j < t\}$
and *rv*: $\forall\, j{<}t.\ random\text{-}variable\ borel\ (\lambda\omega.\ Z\ j\ a\ \omega)$
and $\forall\, j{<}t.\ \ Z\text{-}hat\ j\ a\ \omega =\ \ c * (if\ \pi\ j\ b = a\ then\ Z\ j\ a\text{-}star\ b\ else\ 0)$
and *indep-interval-bounded-random-variables M I X-new a-bound b-bound*
and *indep-loc*: *indep-interval-bounded-random-variables M* $\{j.\ j < t\}$
$(\lambda j.\ (\lambda\omega.\ Z\text{-}hat\ j\ a\ \omega))$
$(\lambda j.\ l\text{-}hat)\ (\lambda j.\ u\text{-}hat)$
and *H*: *Hoeffding-ineq M* $\{j.\ j < t\}$
$(\lambda j.\ (\lambda\omega.\ Z\text{-}hat\ j\ a\ \ \omega))$
$(\lambda j.\ l\text{-}hat)\ (\lambda j.\ u\text{-}hat)$
and *sum-integrals-eq*: $(\sum\ j \in \{j.\ j < t\}.\ integral^L\ M\ (\lambda\omega.\ Z\text{-}hat\ j\ a\ \omega)) = \mu\text{-}hat$
and *rewriting*: $prob\ \{\omega \in \Omega.\ (\sum\ j < t.\ Z\text{-}hat\ j\ a\ \omega) - (\sum\ j < t.\ expectation\ (Z\text{-}hat\ j\ a)) \geq\ \varepsilon\} =$
$prob\ \{x \in space\ M.\ (\sum\ j < t.\ Z\text{-}hat\ j\ a\ x) \geq (\sum\ j < t.\ expectation\ (Z\text{-}hat\ j\ a)) + \varepsilon\}$
shows $prob\ \{\omega \in \Omega.\ (\sum\ j < n.\ Z\text{-}hat\ j\ a\ \omega) \geq \mu\text{-}hat + \varepsilon\}\ \leq \delta$
⟨*proof*⟩

end
end
theory *Discrete-UCB-Step3*
imports *Discrete-UCB-Step2*

begin

locale *bandit-policy* = *bandit-problem* + *prob-space* +
fixes
Ω :: *'b set*
and $\mathcal{F}$ :: *'b set set*
and π :: *nat* $\Rightarrow$ *'b* $\Rightarrow$ *'a*
and *N-n* :: *nat* $\Rightarrow$ *'a* $\Rightarrow$ *'b* $\Rightarrow$ *nat*
and *Z* :: *nat* $\Rightarrow$ *'a* $\Rightarrow$ *'b* $\Rightarrow$ *real*
and δ :: *real*
and *q* :: *real*
assumes *fin-A*: *finite A*
and $\omega \in \Omega$
and *a-in-A*: $a \in A$
and *measurable-policy*: $\forall\, t.\ \pi\ t \in$ *measurable M* (*count-space A*)
and *N-n-def*: $\forall\, n\ a\ \omega.$ *N-n n a* ω = *card* $\{t \in \{0..<n\}.\ \pi\ (t+1)\ \omega = a\}$
and *count-assm-pointwise*: $\forall\, n\ \omega.\ (\sum a \in A.$ *real* (*N-n n a* $\omega)) =$ *real n*
and *delta-pos*: $0 < \delta$
and *delta-less1*: $\delta < 1$
and *q-pos*: $q > 0$

begin

definition *sample-mean-Z* :: *nat* $\Rightarrow$ *'a* $\Rightarrow$ *'b* $\Rightarrow$ *real* **where**
sample-mean-Z n a ω = (*1 / real n*) $*$ ($\sum i<n.$ *Z i a* ω)

definition *M-fun* :: *nat* $\Rightarrow$ *'a* $\Rightarrow$ *'b* $\Rightarrow$ *real* **where**
M-fun t a ω = (*if N-n* (*t+1*) *a* ω = *0 then 0*
else ($\sum$ $s < t.$ (*if* π *s* ω = *a then Z s a* ω *else 0*)) / *real* (*N-n t a* ω))

definition *U* :: *nat* $\Rightarrow$ *'a* $\Rightarrow$ *'b* $\Rightarrow$ *real* **where**
U t a ω = *M-fun t a* ω + *q* $*$ *sqrt* (*ln* (*1 / δ*) / (*2* $*$ *real* (*max 1* (*N-n t a* ω))))

definition *A-t-plus-1* :: *nat* $\Rightarrow$ *'b* $\Rightarrow$ *'a* **where**
A-t-plus-1 t ω = (*SOME a.* $a \in A \wedge (\forall\, a'.\ a' \in A \longrightarrow U\ t\ a\ \omega \geq U\ t\ a'\ \omega)$)

definition *prob-eq-Ex* :: *'b set* $\Rightarrow$ *bool* **where**
prob-eq-Ex E $\equiv$ *prob E* = *expectation* ($\lambda\omega.$ *indicator E* ω)

theorem *proposition-15-7*:
assumes

a-in-A: $a \in A$
and $\omega \in \Omega$
and *subopt-gap*: $\Delta\ a > 0$
and *a-not-opt*: $\exists\ a' \in A.\ \Delta\ a' > 0$
and *delta-a*: $\forall\ a \in A.\ \Delta\ a = \mu\ a\text{-}star - \mu\ a$
and $k \leq n$
and *from-UCB-step1*: $\forall\ a \in A.\ expectation\ (\lambda\omega.\ real\ (N\text{-}n\ n\ a\ \omega)) \leq$
$s\ n + expectation\ (\lambda\omega.\ (\sum\ t = k..<n.\ if\ \pi\ (t+1)\ \omega = a \wedge s\ t \leq real\ (N\text{-}n\ t\ a\ \omega)\ then\ 1\ else\ 0))$
and *from-UCB-step2*: $\forall\ a \in A.\ \forall\ t \in \{k..<n\}.\ prob\ (\{\omega \in \Omega.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a\} \cap$
$\{\omega \in \Omega.\ N\text{-}n\ t\ a\ \omega \geq (2 * ln\ (1\ /\ \delta))\ /\ (\Delta\ a)\hat{}2\}) \leq 2 * \delta$
and *eps-pos*: $\varepsilon > 0$
and *t-gt0*: $\forall\ t \in \{k..<n\}.\ t > 0$
and *choice-delta*: $\forall\ t \in \{k..<n\}.\ \delta = 1\ /\ (real\ t\ powr\ \varepsilon)$
and *s-form*: $\forall\ a \in A.\ \forall\ u.\ s\ u = (2 * \varepsilon * ln\ (real\ u))\ /\ ((\Delta\ a)\hat{}2)$
and *subset-meas*:$\forall\ a \in A.\ \forall t \in \{k..<n\}.\ \forall \omega \in \Omega.\ \{\omega.\ \pi\ (t+1)\ \omega = a \wedge 2 * \varepsilon * ln\ (real\ t)/(\Delta\ a)\hat{}2 \leq N\text{-}n\ t\ a\ \omega\} \subseteq \Omega$
and *prob-eq-E-assm*: $\forall\ a \in A.\ \forall t \in \{k..<n\}.\ prob\ \{\omega.\ \pi\ (Suc\ t)\ \omega = a \wedge 2 * \varepsilon * ln\ (real\ t)\ /\ (\Delta\ a)\hat{}2 \leq real\ (N\text{-}n\ t\ a\ \omega)\} =$
$prob\ (\{\omega.\ \pi\ (Suc\ t)\ \omega = a \wedge 2 * \varepsilon * ln\ (real\ t)\ /\ (\Delta\ a)\hat{}2 \leq real\ (N\text{-}n\ t\ a\ \omega)\} \cap space\ M)$
and *finiteness*: $\forall t{\in}\{k..<n\}.\ \forall a{\in}A.\ emeasure\ M\ \{\omega.\ \pi\ (t+1)\ \omega = a \wedge 2{*}\varepsilon{*}ln(real\ t)/(\Delta\ a)\hat{}2 \leq real\ (N\text{-}n\ t\ a\ \omega)\} < \infty$
and *measurable-set*: $\forall t{\in}\{k..<n\}.\ \forall a{\in}A.\ \{\omega.\ \pi\ (t+1)\ \omega = a \wedge 2{*}\varepsilon{*}ln(real\ t)/(\Delta\ a)\hat{}2 \leq real\ (N\text{-}n\ t\ a\ \omega)\} \in sets\ M$

and *eq-sets-optimum*:
$\forall\ a \in A.\ \forall\ t \in \{k..<n\}.\ \{\omega.\ \pi\ (t+1)\ \omega = a \wedge 2 * \varepsilon * ln\ (real\ t)\ /\ (\Delta\ a)\hat{}2 \leq real\ (N\text{-}n\ t\ a\ \omega)\} =$
$\{\omega \in \Omega.\ A\text{-}t\text{-}plus\text{-}1\ t\ \omega = a\} \cap \{\omega \in \Omega.\ N\text{-}n\ t\ a\ \omega \geq (2 * \varepsilon * ln\ (real\ t))\ /\ (\Delta\ a)\hat{}2\}$

shows
$\forall\ a \in A.\ \ expectation\ (\lambda\omega.\ real\ (N\text{-}n\ n\ a\ \omega)) \leq s\ n + (\sum\ t = k..<n.\ 2\ /\ (real\ t\ powr\ \varepsilon))$

⟨*proof*⟩

theorem *theorem-15-4*:
assumes
a-in-A: $a \in A$
and *finite* A **and** $\forall a \in A.\ integrable\ M\ (\lambda\omega.\ real\ (N\text{-}n\ n\ a\ \omega))$
and *ω-in-Ω*: $\omega \in \Omega$
and *subopt-gap*: $\forall\ a \in A.\ \Delta\ a > 0$
and *a-not-opt*: $\exists\ a' \in A.\ \Delta\ a' > 0$
and *delta-a*: $\forall\ a \in A.\ \Delta\ a = \mu\ a\text{-}star - \mu\ a$
and $k \leq n$
and *n-count-assm-pointwise*: $(\sum a{\in}A.\ real\ (N\text{-}n\ n\ a\ \omega)) = real\ n$

and *expected-regret-prop-15-1*: *expectation* ($\lambda\omega$. *R-n n* ω) = ($\sum a{\in}A$. Δ *a* $*$ *expectation* ($\lambda\omega$. *real* (*N-n n a* ω)))
and *from-UCB-step1*: $\forall$ *a* $\in$ *A*. *expectation* ($\lambda\omega$. *real* (*N-n n a* ω)) $\leq$
s n + *expectation* ($\lambda\omega$. ($\sum$ *t* = *k*..<*n*. *if* π (*t*+*1*) ω = *a* $\wedge$ *s t* $\leq$ *real* (*N-n t a* ω) *then 1 else 0*))
and *from-UCB-step2*: $\forall$ *a* $\in$ *A*. $\forall$ *t* $\in$ {*k*..<*n*}. *prob* ({$\omega \in \Omega$. *A-t-plus-1 t* ω = *a*} $\cap$
{$\omega \in \Omega$. *N-n t a* $\omega \geq$ (*2* $*$ *ln* (*1* / δ)) / (Δ *a*)^*2*}) $\leq$ *2* $*$ δ
and *eps-pos*: ε > *0*
and *t-gt0*: $\forall$ *t* $\in$ {*k*..<*n*}. *t* > *0*
and *choice-delta*: $\forall$ *t* $\in$ {*k*..<*n*}. δ = *1* / (*real t powr* ε)
and *s-form*: $\forall$ *a* $\in$ *A*. $\forall$ *u*. *s u* = (*2* $*$ ε $*$ *ln* (*real u*)) / ((Δ *a*)^*2*)
and *subset-meas*:$\forall$ *a* $\in$ *A*. $\forall$ *t* $\in$ {*k*..<*n*}. $\forall \omega \in \Omega$. {$\omega$. π (*t*+*1*) ω = *a* $\wedge$ *2* $*$ ε $*$ *ln* (*real t*)/(Δ *a*)^*2* $\leq$ *N-n t a* ω} $\subseteq \Omega$
and *prob-eq-E-assm*: $\forall$ *a* $\in$ *A*. $\forall$ *t* $\in$ {*k*..<*n*}. *prob* {ω. π (*Suc t*) ω = *a* $\wedge$ *2* $*$ ε $*$ *ln* (*real t*) / (Δ *a*)^*2* $\leq$ *real* (*N-n t a* ω)} =
prob ({ω. π (*Suc t*) ω = *a* $\wedge$ *2* $*$ ε $*$ *ln* (*real t*) / (Δ *a*)^*2* $\leq$ *real* (*N-n t a* ω)} $\cap$ *space M*)
and *finiteness*: $\forall$ *t*$\in${*k*..<*n*}. $\forall$ *a*$\in$*A*. *emeasure M* {ω. π (*t*+*1*) ω = *a* $\wedge$ *2*$*$$\varepsilon$$*$*ln*(*real t*)/($\Delta$ *a*)^*2* $\leq$ *real* (*N-n t a* ω)} < ∞
and *measurable-set*: $\forall$ *t*$\in${*k*..<*n*}. $\forall$ *a*$\in$*A*. {ω. π (*t*+*1*) ω = *a* $\wedge$ *2*$*$$\varepsilon$$*$*ln*(*real t*)/($\Delta$ *a*)^*2* $\leq$ *real* (*N-n t a* ω)} $\in$ *sets M*

and *eq-sets-optimum*:
$\forall$ *a* $\in$ *A*. $\forall$ *t* $\in$ {*k*..<*n*}. {ω. π (*t*+*1*) ω = *a* $\wedge$ *2* $*$ ε $*$ *ln* (*real t*) / (Δ *a*)^*2* $\leq$ *real* (*N-n t a* ω)} =
{$\omega \in \Omega$. *A-t-plus-1 t* ω = *a*} $\cap$ {$\omega \in \Omega$. *N-n t a* $\omega \geq$ (*2* $*$ ε $*$ *ln* (*real t*)) / (Δ *a*)^*2*}
and *assms-lin-expect*: $\forall$ *a* $\in$ *A*. *expectation* ($\lambda\omega$. $\sum$ *t* = *k*..<*n*.
(*if* π (*t*+*1*) ω = *a* $\wedge$ *2* $*$ ε $*$ *ln* (*real t*) / (Δ *a*)^*2* $\leq$ *real* (*N-n t a* ω) *then 1 else 0*)) =
($\sum$ *t* = *k*..<*n*. *expectation* ($\lambda\omega$. *indicat-real* {ω. π (*t*+*1*) ω = *a* $\wedge$ *2* $*$ ε $*$ *ln* (*real t*) / (Δ *a*)^*2* $\leq$ *real* (*N-n t a* ω)} ω))
and *mono-sum-sets*:
($\forall$ *a* $\in$ *A*. Δ *a* $*$ *expectation* ($\lambda\omega$. *real* (*N-n n a* ω)) $\leq$ Δ *a* $*$ (*s n* + ($\sum$ *t* = *k*..<*n*. *2* / (*real t powr* ε))))
$\Longrightarrow$ ($\sum a \in A$. Δ *a* $*$ *expectation* ($\lambda\omega$. *real* (*N-n n a* ω))) $\leq$
($\sum a \in A$. Δ *a* $*$ (*s n* + ($\sum$ *t* = *k*..<*n*. *2* / (*real t powr* ε))))

shows *expectation* ($\lambda\omega$. *R-n n* ω) $\leq$ ($\sum a{\in}A$. Δ *a* $*$ ((*2* $*$ ε $*$ *ln* (*real n*)) / ((Δ *a*)^*2*)+ ($\sum$ *t* = *k*..<*n*. *2* / (*real t powr* ε))))
⟨*proof*⟩

end
end

References

[1] R. Rebeschini. Lecture 15: Stochastic multi-armed bandit problem and algorithms. 2022. Available at https://www.stats.ox.ac.uk/~rebeschi/teaching/AFoL/22/material/lecture15.pdf.