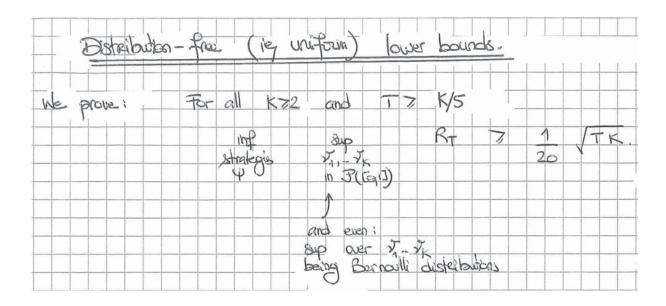
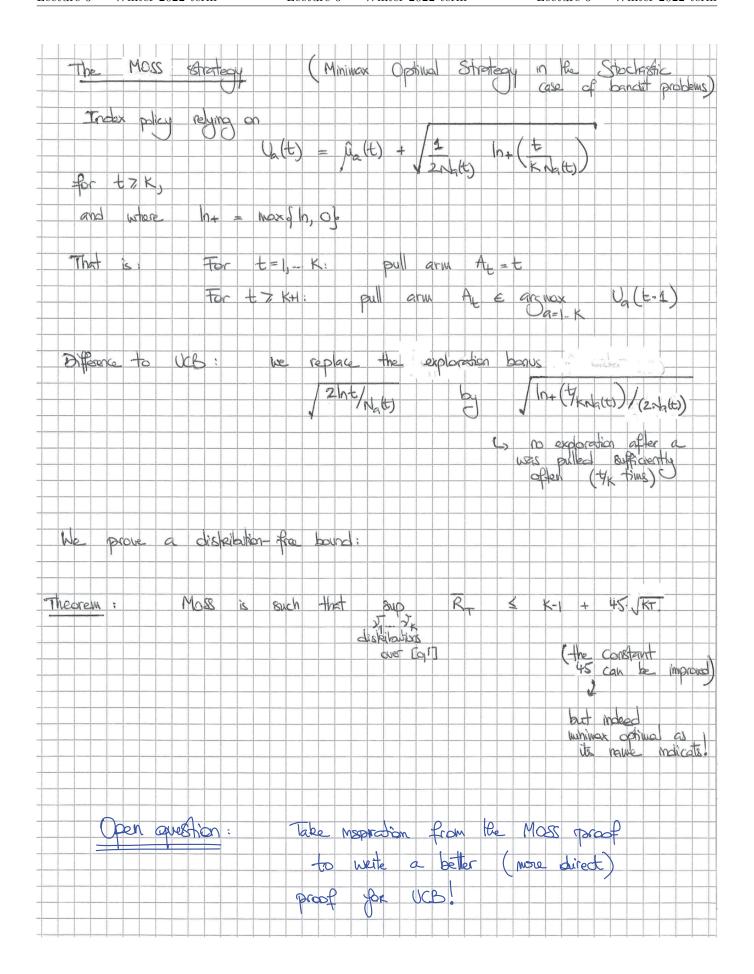
Part 1:  $\sqrt{KT}$  distribution-free regret bounds for stochastic bandits

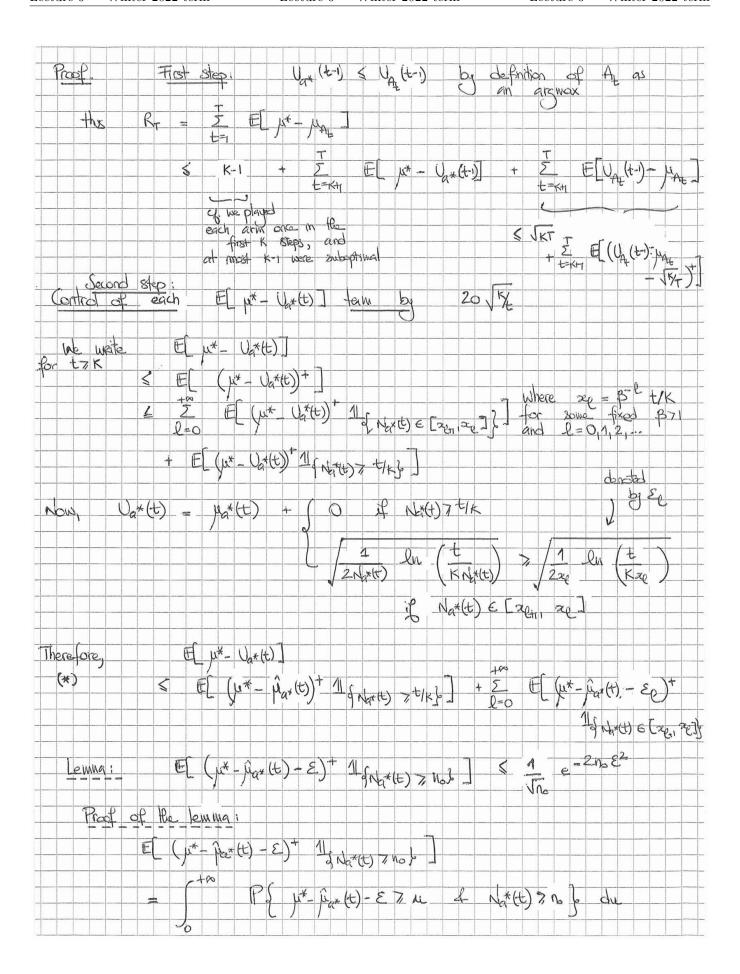
An exercise of the homework is about proving:

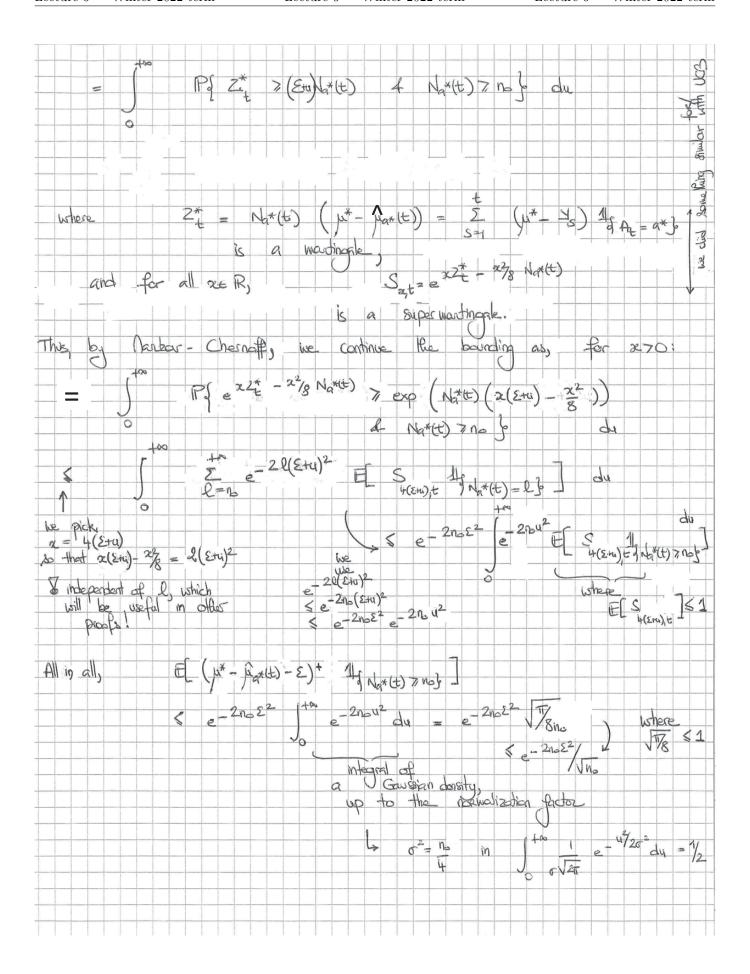


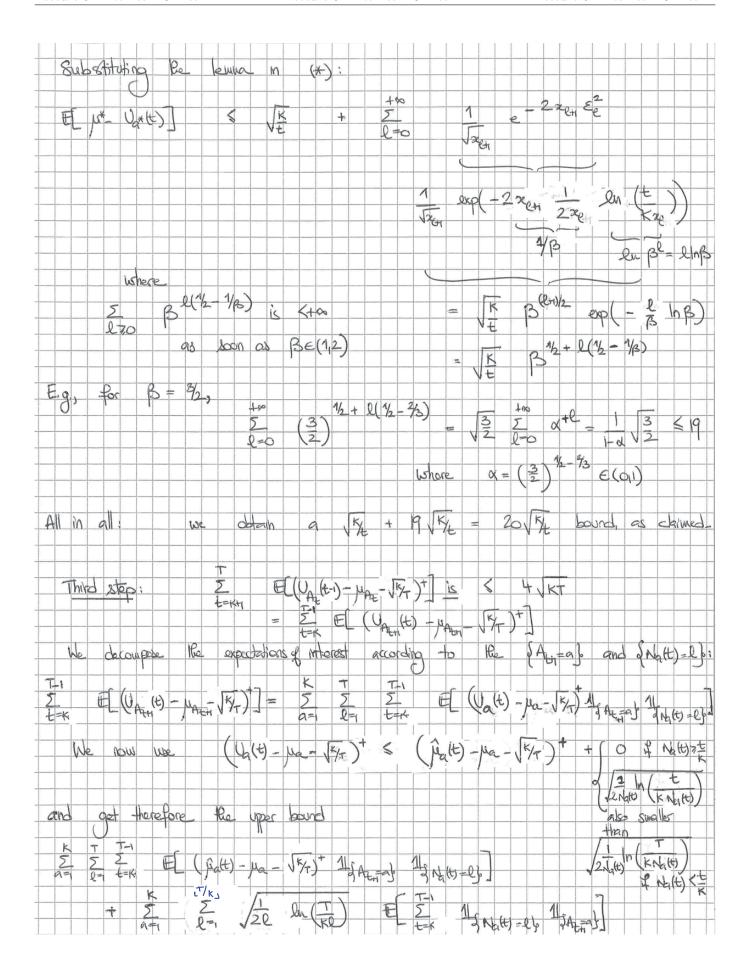
We saw that UCB enjoyed a distribution-free regret bound of order  $\sqrt{TK \ln T}$ , but the  $\sqrt{\ln T}$  is unnecessary. The optimal (minmax) distribution-free regret bound for bounded stochastic bandits is of order  $\sqrt{TK}$ .

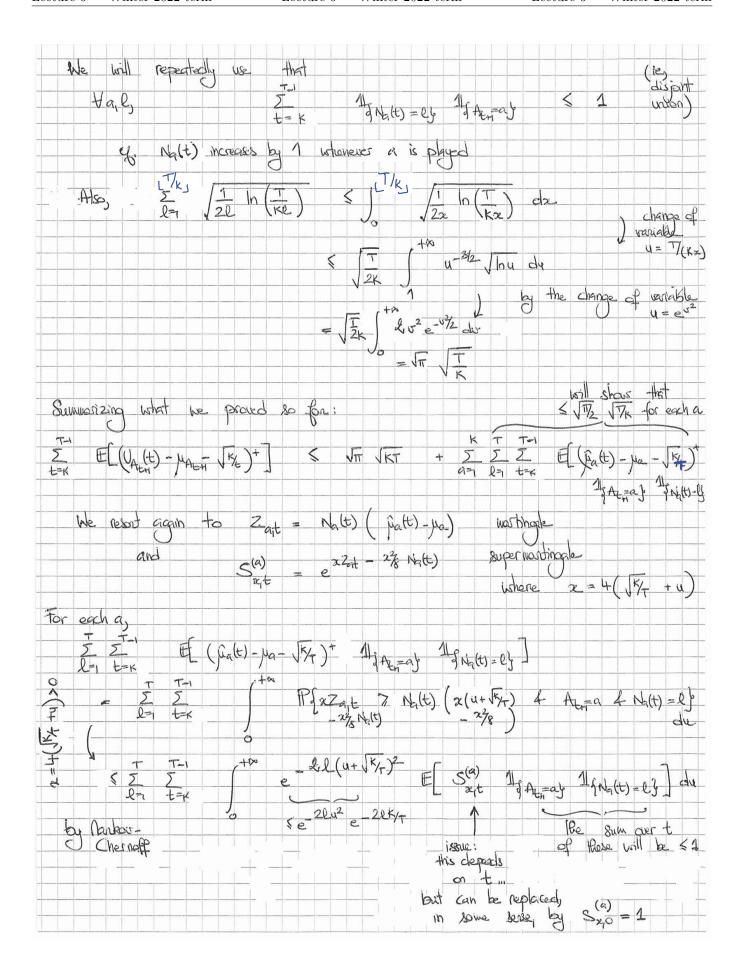
We now discuss an algorithm achieving this optimal order of magnitude; it is called MOSS and is a variation on UCB, with a smaller / more careful exploration bonus.

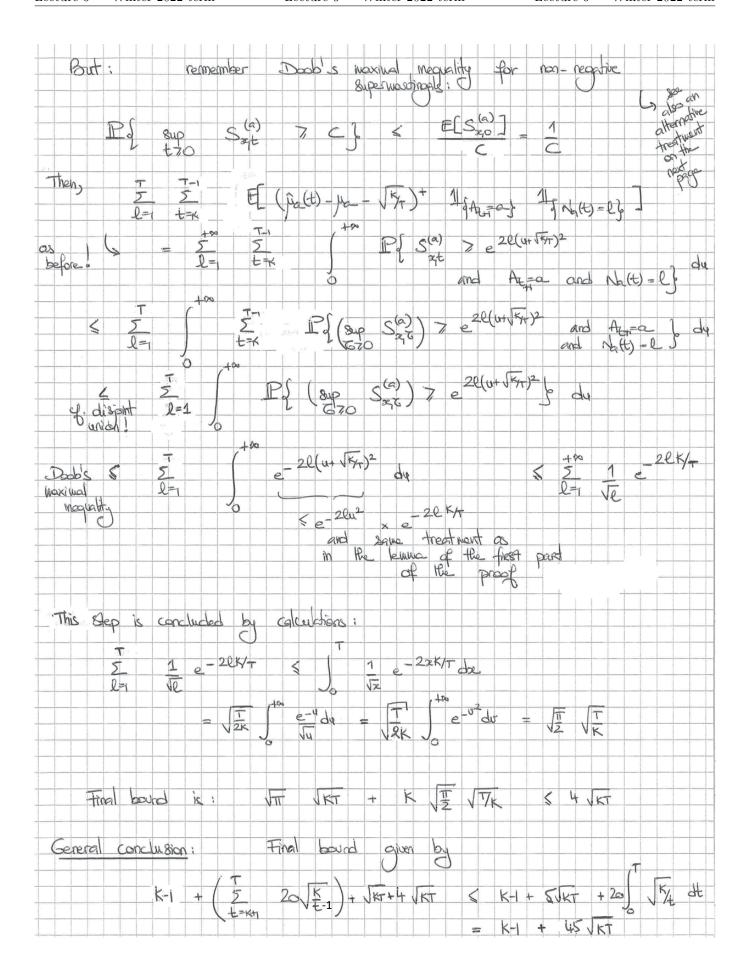


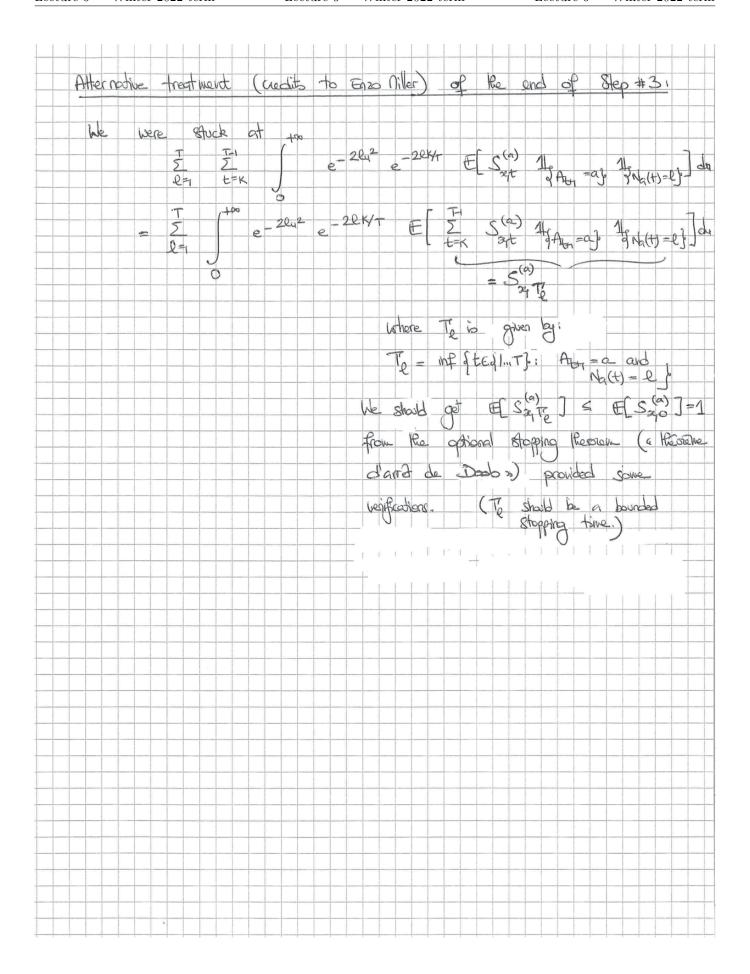












Part 2: Adversarial bandits

	brendits. (Rather Stated in terms of losses ! than remainds!)						
Settingi	At each round t=1,2,						
	1. The apparent and the decision-maker symuthaneously						
	chase lie = (lift) jeji Nje and Ii ~ Pt, where						
	ρ <sub>ε</sub> ε β(ξ), Ny)						
	2. The apponent ats to see pe and It;						
	the decision-maker only observes lite (her own loss).						
Regret:	$R_{T} = \sum_{t=1}^{T} l_{T_{t}, t} - \min_{t=1}^{T} l_{t}$						
vs. Pseudo	record: $R_T = \mathbb{E}\left[\sum_{t=1}^T l_{T_t t}\right] - \min_{j=1,\dots,j} \mathbb{E}\left[\sum_{t=1}^T l_{j t}\right]$						
	Tours de Ration on the Property						
	Lame definition as for stochastic bandits, up to the conversion of loses lift into rewards M-lift Why E? (for a well chosen bound M)  Cf. lift are						
	(for a well chosen bound M) Cf. If are						
	of loses ly into rewards M-lyt Why E?  (for a well chosen bound M)  remakem whichly as  they depend on the pro-						
1 1 5	and in fortular on 1						
we have k	`T						
We have R	`T						
whe actually at RT will be	her shoot for high-probability bounds on RT, but studying						
We actually rat	her shoot for high probability bounds on RT, but studying						
We actually rath	her shoot for high probability bounds on RT, but studying						
We actually rat	her shoot for high probability bounds on RT, but studying						
We actually rath	her shoot for high probability bounds on RT, but studying or good warm up!						
We actually rath	her shoot for high-probability bounds on RT, but studying  a good warm-up!  In these lecture rate, I'll take N = K as the						
We actually rath	her shoot for high probability bounds on RT, but studying  a good warm up!  In these lecture rate, I'll take N = K as the  number of compenents						



The proof is	based on	the follo	wing lew	wa <b>-</b>		UNBOUNDE
Lemma: The		y weighted	avige	strategy o	n losses	is who we develop a
with nt 1,		7	exp(-n, s	t-1 2 ) / N	- exp(-	ver about por
is such that T S. t=		- min i=1~\	て る 見計	< \frac{\ln\lambda}{\eta\tau}\tau	+ Z + =	7t 5 Pji lit
Proof: We s	aus earlier in		s of lea	Junes Heat	the	EWA
Alfe R,	艺 的 到	- min	T	< \lind + \end{y} +	7 2 5 St	2
where $\hat{s}_t$	= 5 71 24	+ i Dr	は 一方 では 。	ુ- ૧૬ ચુંધ		
We us here	e-3- <	1-2+22 2	4270			
So that	りを対す		ln (1 -	It of Pop Pit	+ 1½2 2	
		ln( Ha) ≤u ∀u7-1	- nt	÷ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	7t <sup>2</sup> 5	PJ Ly
herce the stated	bound.					
Proof of the the	coreun); I	le have po they could	control	on how la	rge Re	ly con be,
e ready to app is such Rat		b will a	remainde	MINN ilis M, coul	teim,	whose Mr
Super-linear.	l			the beginning		

