

# IN-STK 5000 Session 10

## Bandits and Model Interpretation



UNIVERSITETET  
I OSLO

# Multi-Armed Bandits

Set of real distributions  
 $B = \{R_1, \dots, R_K\}$ , each  
associated with rewards delivered  
by one of  $K$  levers. Each action  
generates a reward  $r_t$ .  
 $\mu_k = E(R_k)$ . Regret is defined as

$$\rho = T\mu^* - \sum_{t=1}^T \hat{r}_t.$$



# Multi-Armed Bandits

Relevant for many applications,  
e.g.

- Clinical trials
- Adaptive routing
- Financial portfolio design



# Bandits: Strategies

If action  $a$  chosen, and reward  $r$  received, update

$\alpha_a \rightarrow \alpha_a + r$ ,  $\beta_a \rightarrow \beta_a + (1 - r)$ . Probability model  $\hat{\theta} = \alpha / (\alpha + \beta)$ .

- Greedy
  - Explore phase, then exploit
- Epsilon-greedy
  - Choose random action with probability  $\epsilon$
- Probability matching (e.g. Thompson Sampling)
  - Choose action according to probability model

# Bootstrapping

Resample sample data to allow inference of statistical properties of original distribution.

1. Start with  $N$  samples  $\hat{X}$ .
2. Draw  $M$  bootstrap samples  $\hat{B}_i, i = 1, \dots, M$  of size  $N$ , with replacement.
3. Obtain  $M$  values for a statistic  $f_i = F(\hat{B}_i)$  of interest.
4. Use those to e.g. estimate confidence intervals.

