Week 5 - Lecture 27

Strong Law of large numbers

— If we take sufficient samples then we can estimate the true mean

Assumption
↪ The samples drawn should be independent from past pulls of the arm or pulls from other arms. (i.i.d.)

We need to know how many samples to pick before getting a good estimate of the arm, how quickly can I reach this value, since we have limited pulls.

Concentration of Measures

Let $X_1, X_2, \ldots X_n$ be i.i.d. random variables with mean $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathrm{Var}[X_i]$

common mean   common variance

Natural estimator :

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \mathbb{E}[\hat{\mu}] = \mu \quad \begin{bmatrix}\text{unbiased} \\ \text{estimator}\end{bmatrix}$$

This is true for any n

$$\mathrm{Var}[\hat{\mu}] = \mathbb{E}\left[(\hat{\mu}-\mu)^2\right]$$
$$= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right]$$

$$= \frac{1}{n^2} \mathbb{E}\left[\left(\sum_{i=1}^{n} (X_i - \mu) \right)^2 \right]$$

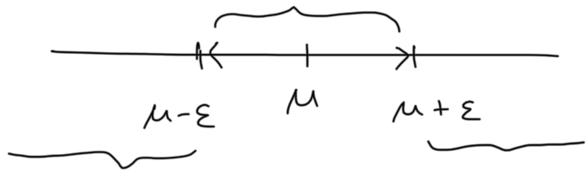
$$= \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu)^2 \right]$$

$$Var[\hat{\mu}] = \frac{\sigma^2}{n}$$ ← as n increases $Var[\hat{\mu}]$ decreases, error reduces with increase in number of samples

Tail Probabilities

$$\underbrace{Pr(\hat{\mu} \geq \mu + \varepsilon)}_{\text{Upper tail probability}} \quad \text{or} \quad \underbrace{P(\hat{\mu} \leq \mu - \varepsilon)}_{\text{Lower tail probability}}$$

$\hat{\mu}$ should be here



Bad regions for $\hat{\mu}$

$Pr\{ |\hat{\mu} - \mu| \geq \varepsilon \}$ → two-sided tail probability

We want the tail probabilities to be small and figure out how it depends on n

$Pr\{ |\hat{\mu} - \mu| < \varepsilon \}$ → how many samples do we need to collect so that the estimate lies in the epsilon region.

Thus we need to bound these.

Inequalities

1) **Markov inequality**

$$\Pr\{|x| \geq \varepsilon\} \leq \frac{\mathbb{E}[|x|]}{\varepsilon}$$

2) **Chebychev inequality**

$$\Pr\{|x - \mathbb{E}[x]| \geq \varepsilon\} \leq \frac{Var(x)}{\varepsilon^2}$$

Now,

$$x = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$Var(\hat{\mu}) = \frac{\sigma^2}{n\varepsilon^2}$$

← with more samples the difference between μ and $\mathbb{E}[\mu]$ will reduce

Now for a fixed n, if ε is small, then the $\frac{\sigma^2}{n\varepsilon^2}$ will become larger, then Probability will be smaller

---

Central Limit Theorem

$x_1, x_2 \ldots x_n \rightarrow$ i.i.d. sequence

$$\frac{\sum_{i=1}^{n} x_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0,1)$$

If $n \to \infty$ then average will converge to normal distribution asymptotically

---

$$Pr\{\hat{\mu} - \mu \geq \varepsilon\}$$

$$= Pr\left\{\frac{1}{n}\sum x_i - \mu \geq \varepsilon\right\}$$

$$= Pr\left\{\frac{\sum x_i - n\mu}{\sqrt{n\sigma^2}} \geq \frac{\varepsilon n}{\sqrt{n\sigma^2}}\right\}$$

$$= Pr\left\{\underbrace{\frac{\sum x_i - n\mu}{\sqrt{n\sigma^2}}} \geq \frac{\varepsilon\sqrt{n}}{\sqrt{\sigma^2}}\right\}$$

If n is sufficiently large the this will converge to $N(0,1)$ (CLT)

$$\approx \int_{\frac{\varepsilon\sqrt{n}}{\sqrt{\sigma^2}}}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-x^2}{2}\right\}}_{\text{PDF for } N(0,1)} dx$$

This integration does not have closed form, thus need

$\left[\text{Let } u = \frac{\varepsilon\sqrt{n}}{\sqrt{\sigma^2}}\right.$

to look at bounds

$$\leq \frac{1}{\sqrt{2\pi}} \int_u \frac{x}{u} \exp\left\{\frac{-x^2}{2}\right\} dx$$

$$= \sqrt{\frac{\sigma^2}{2\pi n \varepsilon^2}} \exp\left(\frac{-n\varepsilon^2}{2\sigma^2}\right)$$

Thus $\Pr\{\hat{\mu} - \mu \geq \varepsilon\}$ is <u>decaying exponentially with n</u>, (with CLT),

with Chebychev it was inversely proportional to n.

But assuming that n is sufficiently large.