# Probability and Bayes' Review

Discretely...

Marginal Distributions $p(A), p(B)$

Joint Distributions $p(A,B) = p(A)p(B)$ if $A, B$ are independent events.

Conditional Distributions $p(A|B), p(B|A)$

$$p(A) = \sum_B p(A,B), \quad p(B) = \sum_A p(A)$$

$$p(A|B) = \frac{p(A,B)}{p(B)} = \frac{p(A,B)}{\sum_A p(A,B)}, \quad p(B|A) = \frac{p(A,B)}{p(A)} = \frac{p(A,B)}{\sum_B p(A,B)}$$

Given $p(A|B)$, $p(B)$, we can also find $p(B|A)$.

$$p(A|B) = \frac{p(A,B)}{p(B)} \Rightarrow p(A,B) = p(A|B)p(B)$$

Bayes' Rule

$$p(B|A) = \frac{p(A,B)}{p(A)} = \frac{p(A,B)}{\sum_B p(A,B)} = \frac{p(A|B)p(B)}{\sum_B p(A|B)p(B)}$$

Continuously...

Joint Probability Density $p(x,y)$

Marginal Probability Density $p(x) = \int p(x,y)\,dy$, $p(y) = \int p(x,y)\,dx$

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x,y)}{\int p(x,y)\,dx} \qquad p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(x,y)}{\int p(x,y)\,dy}$$

$$= \frac{p(y|x)p(x)}{\int p(y|x)p(x)\,dx} \qquad = \frac{p(x|y)p(y)}{\int p(x|y)p(y)\,dy}$$

Bayes' Rule $\longrightarrow$ we can find $p(x|y)$ given that we know $p(y|x)$ and $p(x)$!

# Maximum Likelihood Estimation

Let $X$ be a discrete r.v. with pmf $p$ (or pdf for continuous case) depending on parameter $\theta$. ← parameter that we are trying to solve for our distribution.

$L(\theta|x) = p_\theta(x) = P_\theta(X=x)$ is the likelihood function, given the outcome $x$ of the r.v. $X$.

For Bernoulli r.v. $X$,

$L(\theta|x) = p_\theta(x) = \theta^x(1-\theta)^{1-x}$ s.t. $P_\theta(X=1)=\theta$ ; $P_\theta(X=0) = (1-\theta)$
$= 1 - P_\theta(X=1)$

Suppose we have collected some data ... (Bernoulli coin tosses) i.i.d.

$data = \{x_1, x_2, x_3, \dots, x_N\}$

Likelihood is: $L(\theta|data) = p_\theta(data) = \prod_{i=1}^{N} p_\theta(x_i) = \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{1-x_i}$

↳ Probability of observing the data that we observe assuming each coin toss was i.i.d

e.g. $L(\theta \mid x_1=1, x_2=0, x_3=1) = \theta^1(1-\theta)^1\theta^1$

We want to find the value of $\theta$ that maximizes $L(\theta|data)$!

i.e. the value of $\theta$ that makes the data that we collect most probable.

To Maximize $L(\theta|data)$, we want to take its derivative wrt. $\theta$. s.t. $\frac{dL}{d\theta} = 0$

$\hat{\theta} = \arg\max_\theta L(\theta)$.

Note that argmax is the operation that finds the argument that gives the max value from function $L(\theta)$.

Most of the time it is better to take the log of the likelihood before differentiating,

- and it usually leads to a simpler expression for the derivative that is easier to set to 0 and solve.
- since log is a monotonous function, i.e. whatever maximizes $L$ also maximizes $\log L$.

$l(\theta) = L(\theta) = \log \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{1-x_i} = \sum_{i=1}^{N} \{x_i \log\theta + (1-x_i)\log(1-\theta)\}$.

Set $\frac{dl}{d\theta} = \frac{1}{\theta}\sum_{i=1}^{N} x_i - \frac{1}{1-\theta}\sum_{i=1}^{N}(1-x_i) = 0$

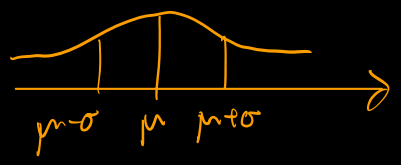
Solve for $\theta$ $(1-\theta)\sum_{i=1}^{N} x_i = \theta\sum_{i=1}^{N}(1-x_i)$.

$\sum_{i=1}^{N} x_i - \theta\sum_{i=1}^{N} x_i = N\theta - \theta\sum_{i=1}^{N} x_i$ $\Rightarrow$ $\theta = \frac{1}{N}\sum_{i=1}^{N} x_i$

Note that for Binomial dist., $P(x=k) = \binom{N}{k}\theta^k(1-\theta)^{N-k}$

For Gaussian (Normal) r.v.'s

$\text{data} = \{x_1, x_2, \dots, x_N\}$ i.i.d.




$$L(\theta \mid \text{data}) = \prod_{i=1}^{N} p_\theta(x_i), \quad \text{where } \theta = \{\mu, \sigma^2\}$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

Goal: Find parameters that best describe data collected.
These values will maximizes the likelihood.

$$\hat{\mu}, \hat{\sigma}^2 = \arg\max_{\mu,\sigma^2} L(\mu, \sigma^2 \mid \text{data})$$

$$l(\mu, \sigma^2) = \log L(\mu, \sigma^2) = \log \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

$$= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] = \sum_{i=1}^{N}\left[\log \frac{1}{\sqrt{2\pi\sigma^2}} + \log e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}\right]$$

$$= \sum_{i=1}^{N}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] = \underbrace{-\frac{1}{2}\log(2\pi\sigma^2)\sum_{i=1}^{N}(1)}_{\text{constant}} - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)\frac{1}{\sigma}$$

Set $\frac{\partial L}{\partial \mu} = 0$, $\quad \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu) = 0.$

$$\sum_{i=1}^{N} x_i - N\mu = 0.$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i = f(x_1, x_2, \dots, x_N).$$

$$l(\mu,\sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$l(\mu, v) = -\frac{N}{2}\log(2\pi v) - \frac{1}{2}\frac{1}{v}\sum_{i=1}^{N}(x_i-\mu)^2, \quad \text{where } v=\sigma^2$$

$$\frac{\partial l}{\partial v} = -\frac{N}{2}\frac{1}{v} - \frac{1}{2}(-1)\frac{1}{v^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

Set $\frac{\partial l}{\partial v} = 0$, $N = \frac{1}{v}\sum_{i=1}^{N}(x_i-\mu)^2$

$$v = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{\mu})^2 = g(x_1, x_2, \dots, x_N).$$

Our MLE estimates are also random variables!

If they are random variables, then we can ask:

- What is their distribution and expectation?

$\hat{\mu} \sim ?$ $\quad \mathbb{E}(\hat{\mu}) = ?$

$\hat{\sigma}^2 \sim ?$ $\quad \mathbb{E}(\hat{\sigma}^2) = ?$

Can use characteristics functions to prove:

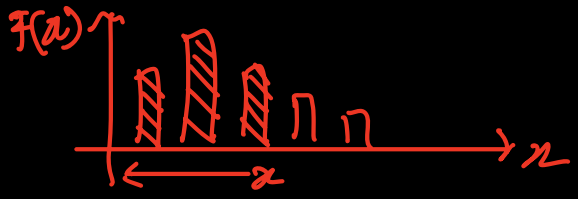
$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$$

$$\mathbb{E}(\hat{\mu}) = \mu$$

$\mathbb{E}(\hat{\sigma}^2) \neq \sigma^2$ ; $\mathbb{E}(\hat{\sigma}^2) = \frac{N-1}{N}\sigma^2 \to \sigma^2$ as $N \to \infty$.

See 'unbiased estimate of the covariance matrix' for full derivation

# CDF and Percentiles




## for Discrete rvs:

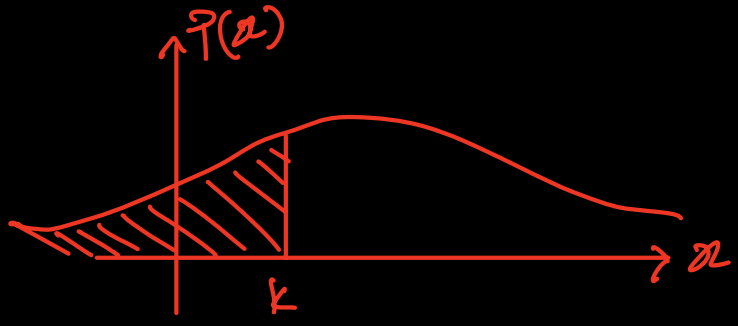
CDF: $F(x) = P(X \leq x) = \sum_{k=-\infty}^{x} p(k)$, where $p(k) = \text{Prob}(X = k)$

## for Continuous rvs:

CDF: $F(k) = P(X \leq k) = \int_{-\infty}^{k} f(t)\,dt$

t is a dummy variable, and disappears after integration.

PDF: $f(x) = \frac{dF(x)}{dx}$




## Inverse CDF (percentile function)

$F^{-1}(p)$ — What value of $x$ would yield a CDF value of $p$?

Eg. Heights — $\mu = 170$ cm, 7cm

$F^{-1}(0.95, \mu = 170, \sigma = 7) = 181.5$

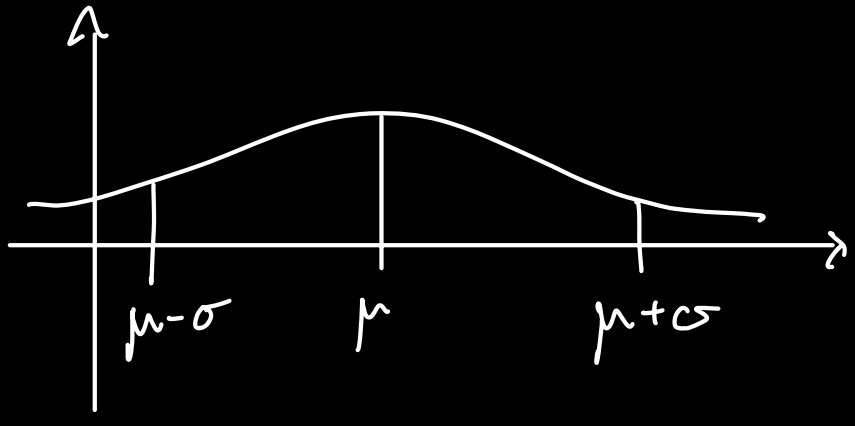
$F(160, \mu = 170, \sigma = 7) = 0.08$

# Traditional A/B Testing.

all about using point estimates to model a population's distribution

Using height as an example, suppose that height has a normal distribution,

$$X \sim N(\mu, \sigma^2)$$




**Note that:**

In probability theory, CLT establishes that in many situations for i.i.d samples, the standardized sample mean tends towards the standard normal distribution even if the original variables are not normally distributed.

We can say that we are more confident that the mean of our sample is representative of the entire population if:

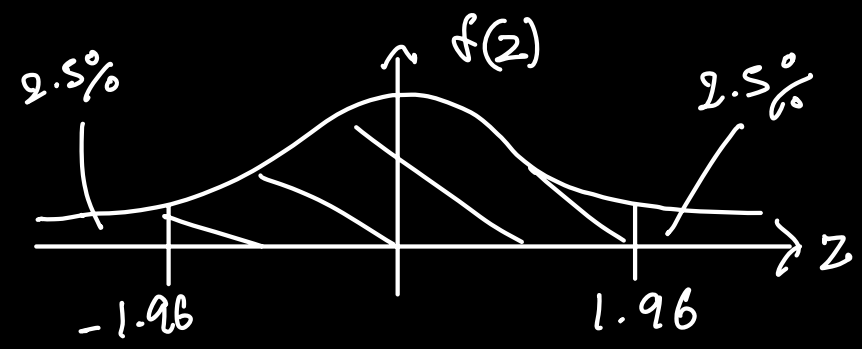
$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$$
$$\left(E(\hat{\mu}),\ Var(\hat{\mu})\right)$$

1. we have a large sample
2. the sample variance is small

In statistics, we typically want to standardize our data so we can easily calculate the probability of certain values occurring in our distribution; or to compare data sets with different means and standard deviations.

i.e. $x \sim N(\mu, \sigma^2) \rightarrow z = \frac{x-\mu}{\sigma} \sim N(0,1)$

For a 95% confidence interval, the lower and upper end points are -1.96 and 1.96.




$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

standardizing a variable.

$$\frac{x-\mu}{\sigma} \sim N(0,1).$$

If $X \sim N(\mu, \sigma^2)$, then $\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$

$$Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}}$$

$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$

$$E(\hat{\mu}) = E\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] = \frac{1}{N}\left[\sum_{i=1}^{N} E(x_i)\right] = \mu$$

$$Var(\hat{\mu}) = E\left[(\hat{\mu} - E(\hat{\mu}))^2\right] = E\left[\left(\frac{1}{N}\sum_{i=1}^{N} x_i - \mu\right)^2\right]$$

$$= \frac{1}{N^2} E\left[\left(\sum_{i=1}^{N} x_i - N\mu\right)^2\right] = \frac{1}{N^2} Var\left(\sum_{i=1}^{N} x_i\right) = \frac{1}{N^2} \cdot N\sigma^2 = \frac{\sigma^2}{N}$$

Note that $\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$ is the unbiased estimator for sample variance (which we should use to replace $\sigma^2$ that is unknown)

for 95% C.I.,

$$-1.96 \leq Z \leq 1.96$$

$$-1.96 \leq \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}} \leq 1.96$$

$$-1.96\left(\frac{\sigma}{\sqrt{N}}\right) - \hat{\mu} \leq -\mu \leq 1.96\left(\frac{\sigma}{\sqrt{N}}\right) - \hat{\mu}$$

$$\hat{\mu} - 1.96\left(\frac{\sigma}{\sqrt{N}}\right) \leq \mu \leq \hat{\mu} + 1.96\left(\frac{\sigma}{\sqrt{N}}\right)$$

$$\hat{\mu} - 1.96\frac{\sigma}{\sqrt{N}} \leq \mu \leq \hat{\mu} + 1.96\frac{\sigma}{\sqrt{N}}$$

replace with $\hat{\sigma}$.

Essentially,

0.95 is the no. of times $\mu$ is contained in the 95% CI out of the total no. of experiments

For a $\gamma$% CI, $\left[\hat{\mu} + \Phi^{-1}\left(\frac{1-\gamma}{2}\right)\frac{\sigma}{\sqrt{N}},\ \hat{\mu} + \Phi^{-1}\left(1 - \frac{1-\gamma}{2}\right)\frac{\sigma}{\sqrt{N}}\right]$

However, we don't know the true variance $\sigma^2$.

Consider a new standardized r.v.
(which is explicitly divided by the estimated sd)

$$t = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{N}} = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{N}} \times \frac{1/\sigma}{1/\sigma} = \left(\frac{\hat{\mu} - \mu}{\sigma/\sqrt{N}}\right) \Big/ \left(\frac{\hat{\sigma}}{\sigma}\right)$$

Consider denominator:

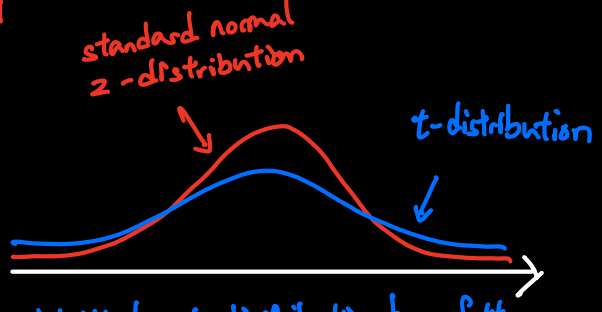
$$\left(\frac{\hat{\sigma}}{\sigma}\right)^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{\sigma}\right)^2$$

$$(N-1)\left(\frac{\hat{\sigma}}{\sigma}\right)^2 \sim \chi^2_{N-1}$$ chi-square distribution with $N-1$ df

Note: $z_i \sim N(0,1) \longrightarrow \sum_{i=1}^{N} (z_i - \bar{z})^2 \sim \chi^2_{N-1}$

$$t = \frac{z}{\sqrt{V/v}} \sim t_v$$

$$\frac{\text{standard normal}}{\sqrt{\text{chi-square} / \text{deg of freedom}}} \sim t_{\text{deg of freedom}}$$




intuitively, t-distribution has fatter tails since we don't know $\sigma$, so we expect the confidence interval to be fatter

t-distribution is parameterized by $(N-1)$ degrees of freedom!

$t_{left} = F^{-1}(0.025;\ df = N-1)$ ; $t_{right} = F^{-1}(0.975;\ df = N-1)$.

$t_{left} \le t \le t_{right}$

$$\hat{\mu} + t_{left}\frac{\hat{\sigma}}{\sqrt{N}} \le \mu \le \hat{\mu} + t_{right}\frac{\hat{\sigma}}{\sqrt{N}}$$

For large values of $N$, t-values $\approx$ z-values

As $N \to \infty$, $\hat{\sigma} \to \sigma \Rightarrow t \to$ Normal

## Z-statistic

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{N}}, \text{ where } \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \; ; \; \hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

and $\sigma$ is replaced with estimator of $\sigma$, $\hat{\sigma}$

$$Z_{left} = \Phi^{-1}(0.025) \; ; \; Z_{right} = \Phi^{-1}(0.975)$$

$$lower = \hat{\mu} + Z_{left} \cdot \frac{\hat{\sigma}}{\sqrt{N}} \; ; \; upper = \hat{\mu} + Z_{right} \cdot \frac{\hat{\sigma}}{\sqrt{N}}$$

## t-statistic

$$t = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{N}} = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{N}} \times \left(\frac{1/\sigma}{1/\sigma}\right) = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{N}} \left(\frac{\hat{\sigma}}{\sigma}\right)$$

$$\Longrightarrow t = \frac{Z}{\sqrt{V/r}} \sim t_r \text{, where } r = N-1 \text{ degrees of freedom}$$

$$t_{left} = \Phi^{-1}(0.025, r = N-1) \; ; \; t_{right} = \Phi^{-1}(0.975, r = N-1)$$

$$lower = \hat{\mu} + t_{left} \cdot \frac{\hat{\sigma}}{\sqrt{N}} \; ; \; upper = \hat{\mu} + t_{right} \cdot \frac{\hat{\sigma}}{\sqrt{N}}$$

# Hypothesis Testing.

- Groups of Data:

  1 → 1-Sample Test  e.g. Avg. daily stock return

  2 → 2-Sample Test  Control vs Treatment Group  e.g. Drug efficacy

- 1-sided vs 2-sided test

Consider 2-sample test for new drug.

(2-sided test)

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

$\leftrightarrow$

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 \neq 0$

(1-sided test)

$H_0: \mu_1 \leq \mu_2$
$H_1: \mu_1 > \mu_2$

→ Output:

Test Statistic  (does it fall in the rejection region?)

✡ P-value

tells us whether the difference is statistically significant (given a significance threshold / level of significance)

Probability to reject the null when the null is true.

How likely that the data observed is to have occurred under the Null Hypothesis.

→ Either reject the null hypothesis,
or fail to reject the null hypothesis

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$$

Consider 1-Sample Test:

For 2-sided test: $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$

Rewrite the null hypothesis

$$H_0: \hat{\mu} \sim N\left(\mu_0, \frac{\sigma^2}{N}\right) \longleftrightarrow H_0: z = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{N}} \sim N(0,1)$$

To calculate p-value,

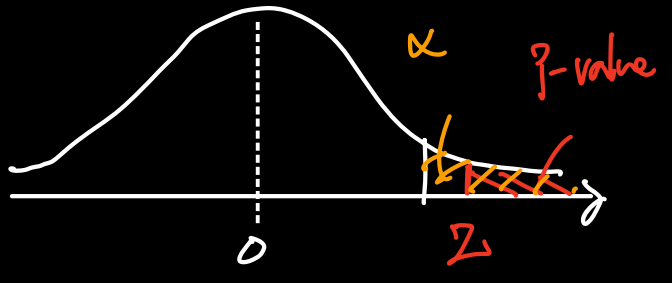
$$p_{left} = \Phi(-|z|) \; ; \; p_{right} = 1 - \Phi(|z|)$$

$$p = p_{left} + p_{right} = p_{left} \times 2 \quad \text{(since distribution is symmetric)}$$

For 1-sided test: $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$

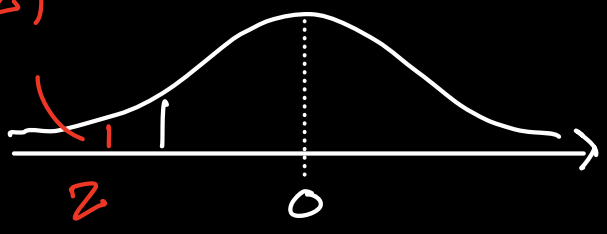
(for the greater than hypothesis)

$$p = 1 - \Phi(z)$$




Similarly, for the lesser than hypothesis,

$$p = \Phi(z)$$



For 2-sample test,

For 2-sided test : $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases} \Leftrightarrow \begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{cases}$

Let $y = \mu_1 - \mu_2$, $\begin{cases} H_0: y = 0 \\ H_1: y \neq 0 \end{cases}$

$$\hat{y} = \hat{\mu}_1 - \hat{\mu}_2 = \frac{\sum_{i=1}^{N_1} x_i}{N_1} - \frac{\sum_{i=1}^{N_2} x_i'}{N_2}$$

$$\begin{aligned} Var(\hat{y}) &= Var\left(\frac{x_1}{N_1}\right) + \dots + Var\left(\frac{x_{N_1}}{N_1}\right) + Var\left(\frac{x_1'}{N_2}\right) + Var\left(\frac{x_{N_2}'}{N_2}\right) \\ &= \frac{\sigma_1^2}{N_1^2} + \dots + \frac{\sigma_1^2}{N_1^2} + \frac{\sigma_2^2}{N_2^2} + \dots + \frac{\sigma_2^2}{N_2^2} \\ &= \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} \end{aligned}$$

$$\sigma_{\hat{Y}} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

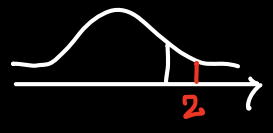
$$Z = \frac{\hat{y} - \mu_0}{\sigma_{\hat{Y}}} = \frac{\hat{y} - \mu_0}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

For P-value, 2-sided test:

$P_{right} = 1 - \Phi(|Z|)$; $P_{left} = \Phi(-|Z|)$; $P = P_{left} + P_{right}$

1-sided test:

$(>)$: $p = 1 - \Phi(Z)$




$(<)$: $p = \Phi(Z)$

# Bayesian A/B Testing

## Explore-Exploit Dilemma

In the frequentist approach, we would decide on how much data to collect beforehand.

i.e. Determine the power and effect size → number of samples needed.

But... prior to experimentation, effect size is UNKNOWN!

(In casino example, effect size would be the difference in win rates between two slot machines.).

Suppose then that we know how many samples to collect...

Even if our prediction of effect size was obviously wrong halfway through sample collection, we must take the experiment to completion.

(Otherwise the experiment would be invalid.)

Exploration (data collection) is a resource-consuming process

However it is necessary when there is no information at hand.

But how much do we need to explore?

Can we know when the right time to exploit the information is?

Ideally, we want a way to best balance explore/exploit.

What algorithms are there to solve this dilemma?

1. Epsilon-Greedy
2. Optimistic Initial Values
3. UCB1 (Upper Confidence Bound)
4. Thompson Sampling (Bayesian Bandit)

These methods overcome some of the awkward problems in traditional A/B testing (since they are adaptive)

# Epsilon-Greedy

means 'short-sighted', using only immediately available information as a heuristic to make a decision.

Problem: Need to balance explore/exploit

simply taking a naive MLE of win rate doesn't work, and this is highly detrimental for exploration/data collection.

we might get lucky or unlucky with a particular bandit, and end up exploiting something suboptimally because the win rate on all the other bandits is just 0 out of 1.

The idea is that we will introduce a small probability ($\varepsilon$) of doing something random (non-greedy)

typically 5% or 10%

PSEUDOCODE:

Greedy

```
while True:
      j = argmax (predicted bandit means)
      x = play bandit j and get reward
      bandits[j].update_mean(x)
```

we select the $j^{th}$ bandit i.e the bandit with the current largest mean

Epsilon-Greedy

```
while True:
      p = random number in [0,1]
      if p < epsilon:
            j = choose a random bandit
      else:
            j = argmax (predicted bandit means)
      x = play bandit j and get reward
      bandits[j].update_mean(x)
```

Additionally, we can consider decaying epsilon

# Iteratively Updating Sample Mean

Sample Mean at time N can be evaluated in constant space and time

$$\bar{X}_N = \frac{1}{N}\left(\sum_{i=1}^{N-1} X_i + X_N\right)$$

$$= \frac{1}{N}\left[(N-1)\bar{X}_{N-1} + X_N\right]$$

$$= \frac{N-1}{N}\bar{X}_{N-1} + \frac{1}{N}X_N$$

$$= (1-\frac{1}{N})\bar{X}_{N-1} + \frac{1}{N}X_N$$

$$= \bar{X}_{N-1} + \frac{1}{N}(X_N - \bar{X}_{N-1})$$

```
class Bandit:
    def __init__(self, p):
        self.p = p
        self.p_estimate = 0
        self.N = 0
    def pull(self):
        return np.random.random() < self.p

    def update(self, x):
        self.N = self.N + 1
        self.p_estimate
          = ((self.N-1) * self.p_estimate + x)
            / self.N
```

## Epsilon Greedy Program Layout.

Define

- No. of Trials — NUM_TRIALS = 10000
- Epsilon — EPS = 0.1 Exploration Probability

Consider Decaying Epsilon!

- BANDIT_PROBABILITIES = [0.2, 0.5, 0.75]

win probability for each bandit (slot machine)

No. of Trials already completed.

### ① Bandit Class (contains 3 attributes, p, p_estimate, N).

p_estimate: current estimated win rate

- pull(self) (pulling arm of slot machine and obtaining a reward)

return True (1) if win False (0) if lose (BERNOULLI CASE)

- update(self, x)

update self.N

update self.p_estimate

## ② Epsilon-Greedy Loop

In each trial,

① explore random bandit with probability Epsilon (EPS)

OR

② exploit bandit with highest estimated win rate

then,

update probability estimates with results from current trial

At end of experiment, we should have logged

① p-estimate for each bandit → which points us to the optimal bandit

② total rewards earned

③ no. of times explored

④ no. of times exploited

⑤ no. of times optimal bandit was selected

IN THE GAUSSIAN CASE, the process is similar

for each bandit, the rewards follow a gaussian distribution

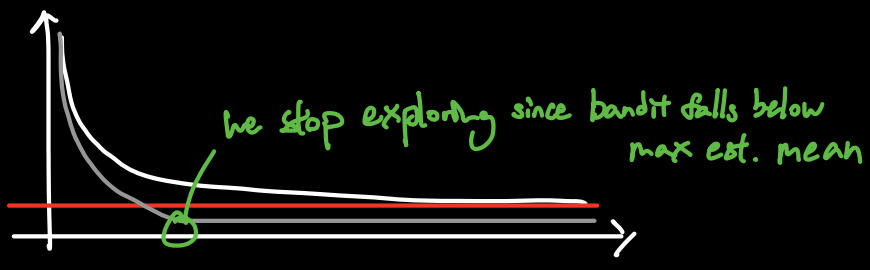
$$X \sim N(M, \sigma^2)$$

## Optimistic Initial Values

Hyperparameter that controls the amount of exploration

- OIV is a simple modification of the purely greedy method.
- No need for Epsilon (random exploration)

Instead of initializing mean estimate 0, give it a very large value.
(overestimate)




## Upper Confidence Bound

Reasonable way to explore / exploit bandits by comparing upper bounds

$$p(\text{sample mean} - \text{true mean} \geq \text{error}) \leq f(\text{error})$$

decreasing error function

e.g $p(\text{sample mean} - \text{true mean} \geq t) \leq 1/t$

Markov Inequality : RHS decreases $\propto 1/t$

Chebyshev Inequality : RHS decreases $\propto 1/t^2$

UCB1 → Hoeffding's Inequality! $P(\bar{X}_n - \mathbb{E}(X) \geq t) \leq e^{-2nt^2}$

$\bar{X}_n$ : sample mean after collecting n samples

$\mathbb{E}(X)$ : expected value of X (true mean)

$\bar{X}_n - \mathbb{E}(X)$ : measurement error of sample mean

$t$ : arbitrary error value

PSEUDOCODE

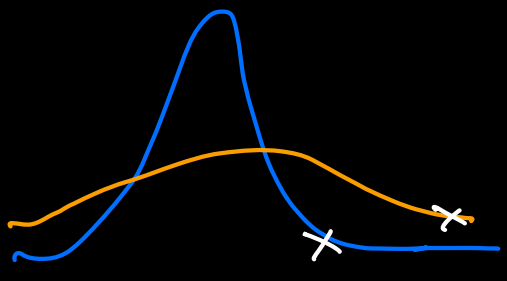
Loop:

$$j = \underset{j}{\operatorname{argmax}} \left( \bar{x}_{nj} + \sqrt{2 \frac{\log N}{n_j}} \right)$$

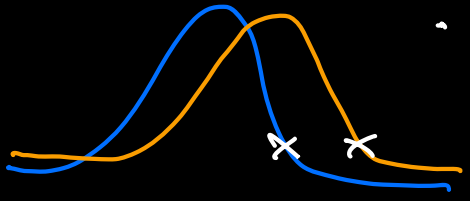
# pull arm j, update bandit j's mean..

Intuitively...

# x shows location of 95% CI



- Not confident → want to explore



- More confident → choose the best bandit

Exploration: bigger when less data

Exploitation: we want to choose bandit with bigger mean

$N$ = total plays made

$n_j$ = plays made on bandit $j$

$$j = \underset{j}{\operatorname{argmax}} \left( \bar{X}_{n_j} + \sqrt{2 \frac{\log N}{n_j}} \right)$$

Heuristic / Hyperparameter

if constant is made bigger, than we will have a larger upper bound, vice versa.

**How does it work?**

If we ignore bandit $j$, then $n_j$ is small.

$\sqrt{2 \frac{\log N}{n_j}}$ ↑ → want to explore bandit $j$ more.

$$\lim_{n \to \infty} \frac{\log n}{n} \to 0$$

Hoeffding's Inequality: $p(\bar{X}_n - \mathbb{E}(X) \geq t) \leq e^{-2nt^2}$

RHS: $p = e^{-2n_j t^2} \iff t = \sqrt{\frac{-\log p}{2n_j}}$

UCB1 as a heuristic, we take the numerator to be $4 \log N$.

i.e. $-\log p = 4 \log N$

$p = N^{-4}$

we want the probability upper bound to drop off as $1/N^4$

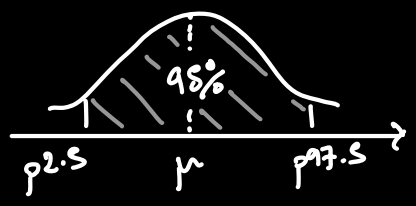
# Bayesian Bandits / Thompson Sampling

CLT → Sum of RVs ~ Normal Distribution

Intuitively,

Small Dataset → Large Confidence Interval, vice versa

Fat → Explore more, Skinny → Explore less




Confidence Intervals are used in the classical Frequentist approach.

In the Bayesian approach, "Everything is a random variable, and has a distribution"

we are more interested in the distribution of the true mean.

In order to determine the distribution of the mean $\theta$.

posterior distribution

$$\underline{p(\theta|X)} = \frac{\underline{p(X|\theta)}\,\underline{p(\theta)}}{\underline{p(X)}}$$

likelihood (probability of data given parameter $\theta$)

prior (distribution of $\theta$ when we don't know anything about $X$)

evidence

Since we are trying to find $p(\theta|X)$ which is a distribution over the r.v. $\theta$;

and $p(X)$ is constant w.r.t $\theta$.

$\Rightarrow$ the above can be interpreted as a proportionality rather than an equation.

(disregarding evidence)

$$p(\theta|X) \propto p(X|\theta)\,p(\theta) \quad \checkmark$$

$$p(\theta|X) = \frac{p(X|\theta)\,p(\theta)}{\int p(X|\theta)\,p(\theta)\,d\theta} \quad \times$$

Difficult to calculate integrals;
Not feasible to run a Monte Carlo simulation to obtain an answer.

Instead, we can rely on Conjugate Priors which are special pairs of distributions that allow us to take advantage of proportionality to ignore the evidence,

where posterior has the same form as the prior.

$p(\theta|X)$ — likelihood function

$p(\theta)$ — conjugate prior

In general, for our common distributions e.g. gaussian,

Gaussian $\propto$ Gaussian x Gaussian

i.e. If we pick the right likelihood and prior, the posterior will be the same kind of distribution as the prior.

For Bernoulli likelihood,

$X = \{x_1, x_2, \ldots, x_N\}$

$p(X|\theta) = \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{(1-x_i)}$, $x_i \sim \text{Bernoulli}(\theta)$

In this case $\theta$ is the Bernoulli parameter

likelihood is expressed as product of pmfs of each $x_i$.

Beta distribution is the conjugate prior for Bernoulli likelihood

$$p(\theta|X) \propto p(X|\theta)\,p(\theta)$$

$$= \left(\prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{(1-x_i)}\right)\left(\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right)$$

In a binary reward situation, the mean of the distribution must be between $[0,1]$

The support of Beta is within the range $[0,1]$

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha)+\Gamma(\beta)}$$

we can think of gamma $\Gamma$ as the generalization of the factorial function of complex numbers

$$p(\theta|X) = \dots$$

$$= \theta^{\alpha - 1 + \sum_{i=1}^{N} x_i} \quad (1-\theta)^{\beta - 1 + \sum_{i=1}^{N}(1-x_i)}$$

we notice that the posterior $p(\theta|X)$ is exactly the same form as a Beta distribution (that of the prior $p(\theta)$)

$$f(\theta; \alpha, \beta) = \text{const} \times \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

we do not have to worry about what the value of this constant is since it's just whatever the value needs to be so that the pdf integrates to 1.

$$p(\theta|X) = \dots = \text{Beta}\left(\alpha + \sum_{i=1}^{N} x_i,\ \beta + N - \sum_{i=1}^{N} x_i\right),$$

if $p(\theta) = \text{Beta}(\alpha, \beta)$

## Choosing $\alpha, \beta$

- $\text{Beta}(1,1)$ is the uniform distribution - $\text{Uniform}[0,1]$ which usually makes a good choice for the Beta prior since if we have no clue what the result may be, we can assign equal probabilities to every possible value.
- If we do have prior knowledge, eg. CTR ≈ 1%/2%, we can consider encoding it into the prior.

## Updating the Posterior in practice (N=1)

- In bandit code, we update the model each time a data point is collected.
- The posterior in one step becomes the prior in the next step.

e.g.

Prior = Beta(1,1)
→ Collect x=1
→ Posterior = Beta(1+1, 1+1-1) = Beta(2,1)

Prior = Beta(2,1)
→ Collect x=1
→ Posterior = Beta(2+1, 1+1-1) = Beta(3,1)

Prior = Beta(3,1)
→ Collect x=0
→ Posterior = Beta(3+0, 1+1-0) = Beta(3,2)

## Picking the Bandit.

Instead of an upper bound, use a sample drawn from posterior.
(Thompson sampling)

By drawing samples from this distribution, we are saying give me a value from this distribution and let that determine which bandit we choose.

So, instead of picking one value (upper bound), we can make use of all possible values under the distribution

(which will become skinnier over time as we become more confident in our beliefs in where the true mean lies.)

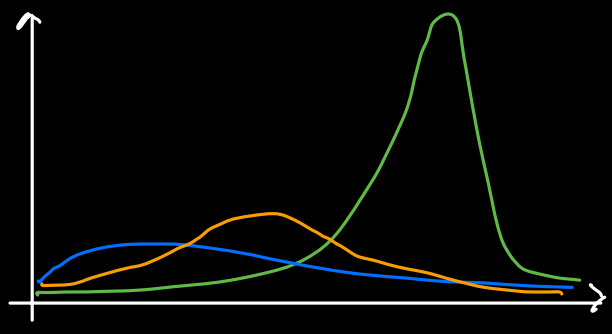
```
PSEUDOCODE:

class Bandit:
      def sample():
            return beta(a,b).sample()

      def update():
            a = ..., b = ...

for n in range(NUM_TRIALS):
      j = argmax(b.sample() for b in bandits)
      x = bandit[j].pull()
      bandit[j].update(x)
```

return a sample of a beta distribution given its current values of $(\alpha, \beta)$



- Bayesian ML gives us tools to get the actual mean posterior.
- In general, computing a posterior from Bayes rule is not easy since it usually involves intractable sums or unsolvable integrals, unless we use conjugate priors.
- Instead, we can use proportionality to prove that the shape of the posterior fits some particular distribution, and from there the normalizing constant can be set so that the integral is 1.
- Thompson Sampling ranks each bandit based on its posterior sample! allowing us to leave suboptimal distributions 'fat' and exploit the optimal bandit more.

# Thompson Sampling on Real-Valued Rewards.

Note that: In Bayesian ML, we usually use precision ($\tau$) (1/variance) instead of variance

Gaussian likelihood: $p(X|\mu,\tau) = \prod_{i=1}^{N} \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x_i-\mu)^2}$

Conjugate Priors:

| Likelihood | Conjugate Prior |
|---|---|
| Unknown Mean, Known Precision | Normal |
| Known Mean, Unknown Precision | Gamma |
| Unknown Mean, Unknown Precision | Normal-Gamma |

Note that: $X \sim N(\mu, \tau^{-1})$ ; $\mu | X \sim N(m, \lambda^{-1})$

i.e. we should not confuse the mean and precision of $\mu$ with the mean and precision of $X$.

AND if we treat both $\mu$ and $\tau$ as r.v.s, then we will have 4 parameters (normal-gamma)

So... How do we calculate the posterior parameters from the prior parameters?

In our example, we shall look at a Gaussian likelihood with fixed precision ($\tau = 1$).

$$p(\mu|X) \propto p(X|\mu)\, p(\mu)$$

$\mu|X \sim N(m, \lambda^{-1})$

- start with prior parameters $m_0$ and $\lambda_0$
- goal is to find $m$ and $\lambda$, the parameters of the posterior distribution of $\mu$ as a function of the data $X$ and the prior parameters $m_0$ and $\lambda_0$

$$m = f_m(X, m_0, \lambda_0), \quad \lambda = f_\lambda(X, m_0, \lambda_0)$$

$$X \sim N(\mu, \tau^{-1}), \quad \mu \sim N(m_0, \lambda_0^{-1}), \quad \mu|X \sim N(m, \lambda^{-1})$$

posterior $\propto$ likelihood $\times$ prior

$$p(\mu|X) \propto p(X|\mu)p(\mu)$$

Note that precision of $X$ has been dropped from the distribution since it is assumed to be fixed.

$$= \left(\prod_{i=1}^{N} \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x_i-\mu)^2}\right)\left(\frac{\lambda_0}{2\pi}\, e^{-\frac{\lambda_0}{2}(\mu-m_0)^2}\right)$$

$$= \left(\left[\sqrt{\frac{\tau}{2\pi}}\right]^N e^{-\frac{\tau}{2}\sum_{i=1}^{N}(x_i-\mu)^2}\right)\left(\sqrt{\frac{\lambda_0}{2\pi}}\, e^{-\frac{\lambda_0}{2}(\mu-m_0)^2}\right)$$

$$\propto \left(e^{-\frac{\tau}{2}\sum_{i=1}^{N}(x_i-\mu)^2}\right)\left(e^{-\frac{\lambda_0}{2}(\mu-m_0)^2}\right)$$

$$= e^{-\frac{\tau}{2}\sum_{i=1}^{N}(x_i-\mu)^2 - \frac{\lambda_0}{2}(\mu-m_0)^2}$$

$$= e^{-\frac{\tau}{2}\sum_{i=1}^{N}(\mu^2 - 2\mu x_i + x_i^2) - \frac{\lambda_0}{2}(\mu^2 - 2\mu m_0 + m_0^2)}$$

$$= \exp\left[-\frac{\tau}{2}\left(N\mu^2 - 2\mu\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i^2\right) - \frac{\lambda_0}{2}\left(\mu^2 - 2\mu m_0 + m_0^2\right)\right]$$

Remember that we don't care about any terms that don't involve $\mu$...

So we can drop these terms.

$$\propto \exp\left[-\frac{\tau}{2}\left(N\mu^2 - 2\mu\sum_{i=1}^{N} x_i\right) - \frac{\lambda_0}{2}\left(\mu^2 - 2\mu m_0\right)\right]$$

$$= \exp\left[-\frac{\tau N + \lambda_0}{2}\mu^2 + \left(\tau\sum_{i=1}^{N} x_i + \lambda_0 m_0\right)\mu\right]$$

We can observe that for the posterior $p(\mu|X)$,

$$p(\mu|X) = \sqrt{\frac{\lambda}{2\pi}}\exp\left[-\frac{\lambda}{2}(\mu - m)^2\right]$$

$$= \sqrt{\frac{\lambda}{2\pi}}\exp\left[-\frac{\lambda}{2}\left(\mu^2 - 2\mu m + m^2\right)\right]$$

$$\propto \exp\left[-\frac{\lambda}{2}\left(\mu^2 - 2m\mu\right)\right]$$

$$= \exp\left[-\frac{\lambda}{2}\mu^2 + m\lambda\mu\right]$$

$$\Rightarrow \lambda = \tau N + \lambda_0$$

$$m\lambda = \tau\sum_{i=1}^{N} x_i + \lambda_0 m_0 \iff m = \frac{1}{\lambda}\left(\tau\sum_{i=1}^{N} x_i + \lambda_0 m_0\right)$$

$$m = \frac{1}{\tau N + \lambda_0}\left(\tau\sum_{i=1}^{N} x_i + \lambda_0 m_0\right)$$

update!

```
self.m = (self.tau * x + self.lambda * self.m) / (self.tau + self.lambda)

self.lambda += self.tau

self.N += 1
```

as we collect more data, the effect of the prior wears off

$\because \lambda \to \infty$, $m \to$ sample mean

Summary

- Thompson Sampling can be used on any likelihood distribution.
- by using the right conjugate prior (refer to wikipedia)
- the general algorithm used remains the same, and all we really need to do is to find the posterior equation!

# Nonstationary Bandits

What happens when our rewards are nonstationary?

ie. what happens when the distribution of our rewards changes over time?

our previous sample mean calculations would no longer be correct.

Most of the time in statistics and signal processing, we are interested in a weak-sense stationarity (where mean and autocovariance doesn't change over time)

In the Sample Mean Calculation, taking the usual sample mean won't work.

$$\bar{X}_N = \frac{1}{N}\left[(N-1)\bar{X}_{N-1} + X_N\right]$$

$$= \frac{N-1}{N}\bar{X}_{N-1} + \frac{1}{N}X_N$$

$$= \left(1 - \frac{1}{N}\right)\bar{X}_{N-1} + \frac{1}{N}X_N$$

$$= \bar{X}_{N-1} + \frac{1}{N}\left(X_N - \bar{X}_{N-1}\right)$$

learning rate

$$\bar{X}_{N-1} + \alpha\left(X_N - \bar{X}_{N-1}\right)$$

Constant learning rate

gives us an exponentially-weighted moving average (EWMA)

Intuitively, as the true mean changes over time, older data becomes less reliable. ⟶ we should put more weight in recent data!

$$\bar{X}_t = \bar{X}_{t-1} + \alpha\left(X_t - \bar{X}_{t-1}\right)$$

Recursively ....

$$\bar{X}_t = (1-\alpha)\bar{X}_{t-1} + \alpha X_t$$

$$= (1-\alpha)^2\bar{X}_{t-2} + (1-\alpha)\alpha X_{t-1} + \alpha X_t$$

$$= \cdots$$

$$= (1-\alpha)^t \bar{X}_0 + \alpha\sum_{k=0}^{t-1}(1-\alpha)^k X_{t-k}$$