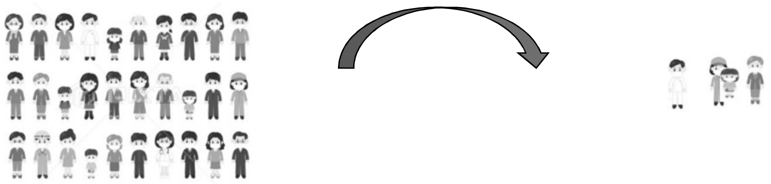
Sampling Distributions

- Sample
  - Population
  - Sample



Sampling Distributions

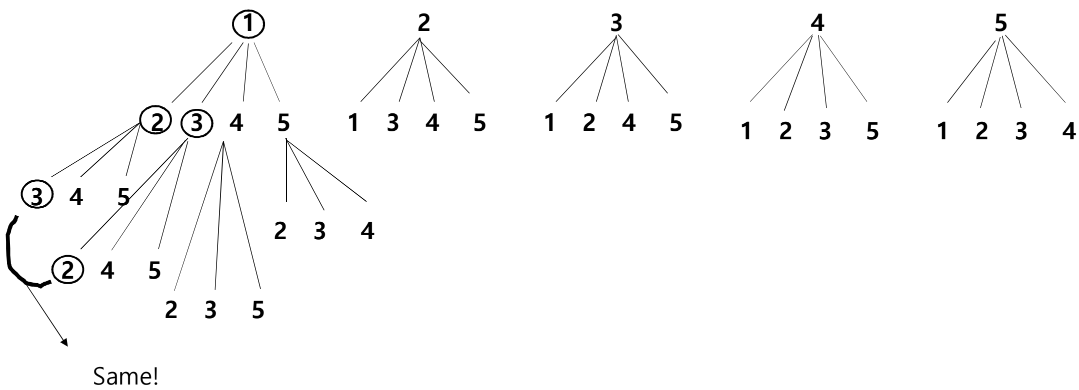
- Sampling
  - Sampling with replacement
    : a sample is drawn from a finite population, and then return to that population after its characteristic has been recorded
  - Sampling without replacement
    : a sample is drawn and its characteristic is recorded.

Sampling Distributions

- Example
  - There are five cards in a pot. The numbers from 1 to 5 are written on the faces of the cards. You randomly drew 3 out of the 5 cards without replacement, and compute the mean values with the selected 3 cards.

Sampling Distributions

- How many kinds of samples you can possibly have?



Sampling Distributions

- Sampling
  - Permutation

$$nP_r = \frac{n!}{(n-r)!}$$

  - Combination

$$nC_r = \frac{n!}{r!\,(n-r)!}$$

Sampling Distributions

- Sample means
  - (1,2,3) ⇨ $\frac{(1+2+3)}{3} = 2.0$
  - (1,2,4) ⇨ $\frac{(1+2+4)}{3} = 2.3$
  - (1,2,5) ⇨ $\frac{(1+2+5)}{3} = 2.6$
  - (1,3,4) ⇨ $\frac{(1+3+4)}{3} = 2.6$
  - (1,3,5) ⇨ $\frac{(1+3+5)}{3} = 3.0$
  - (1,4,5) ⇨ $\frac{(1+4+5)}{3} = 3.3$
  - (2,3,4) ⇨ $\frac{(2+3+4)}{3} = 3.0$
  - (2,3,5) ⇨ $\frac{(2+3+5)}{3} = 3.3$
  - (2,4,5) ⇨ $\frac{(2+4+5)}{3} = 3.6$
  - (3,4,5) ⇨ $\frac{(3+4+5)}{3} = 4.0$

Sampling Distributions

- Distribution of sample means
  - The mean of sample means

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \mu$$

  - The variance of sample means

$$Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

Sampling Distributions

np.random.choice(*x, size, replace, p*)
-. *x* : integer or an 1D-array
-. *size*: integer of a tuple
-. *replace* = True or False
-. *p* : the probability associated with each entry

Sampling Distributions

- LOOP sentence

for i in range(n) :
: Loop n number of times

while condition:
: Loop until the condition is true

Sampling Distributions

- LOOP controls

break:
: Terminate the loop

continue:
: Jump over the current iteration, and go to the next iteration

Probability Distributions

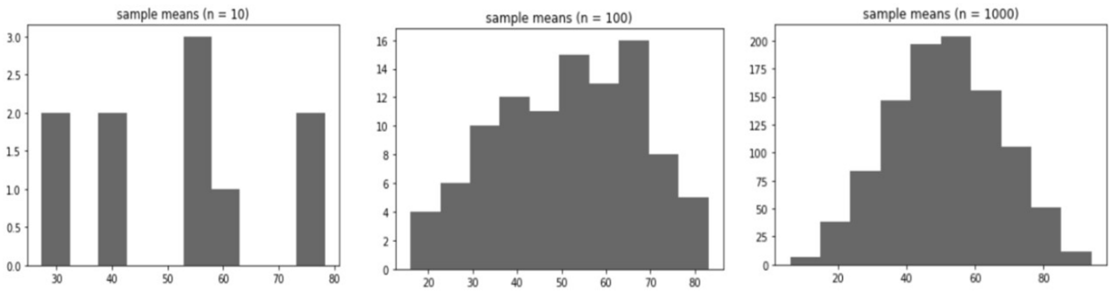
- Practice
  - If 100 cards from 1 to 100 were in a pot instead of 5…

```
In [4]: n = 10
        s = np.arange(1,101)
        s_mean = np.zeros(n)

        for i in range(n):
            x = np.random.choice(s,3,replace=False)
            s_mean[i] = np.mean(x)

        plt.hist(s_mean)
        plt.title('sample means (n = 10)')
```

한양대학교 HANYANG UNIVERSITY

Probability Distributions

- Practice



한양대학교 HANYANG UNIVERSITY

Probability Distributions

- Axiom of Probability
  - $0 \le P(E) \le 1$
  - $P(S) = 1$
  - $P(A \cap B) = \varnothing$, $P(A \cup B) = P(A) + P(B)$

Probability Distributions

- Probability Functions
  - Discrete probability functions

    : The probability function of discrete variable
  - Continuous probability functions

    : The probability function of continuous variable

Probability Distributions

- Bernulli’s events

1 0

  - The outcomes can be either 0 or 1

$$p(x) = p^x \, (1-p)^{1-x}$$

한양대학교 HANYANG UNIVERSITY

Probability Distributions

- Binomial distribution

$$P_K(k) = \binom{n}{k} p^k q^{n-k}, 0 \le k \le n,$$

한양대학교 HANYANG UNIVERSITY

Sampling Distributions

np.random.seed(seed)
-. *seed* : an integer
-. Set the seed value to initialize the random generator

np.random.binomial(n, p, size)
-. *n* : the number of trials
-. *p*: success probability
-. *size*: output shape

한양대학교 HANYANG UNIVERSITY

Sampling Distributions

np.random.random(*size*)
-. *size* : an integer
-. Generate a random number in [0,1) as many as size

np.random.uniform(low, high, size)
-. *lower* : lower boundary
-. *high*: upper boundary
-. *size*: output shape

한양대학교 HANYANG UNIVERSITY

Probability Distributions

- Uniform distribution (a, b)

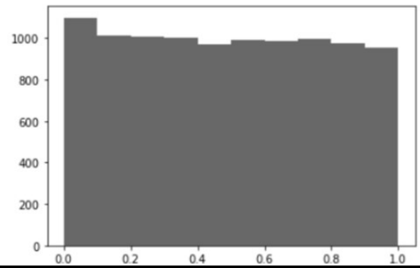
$$f_X(x) = \frac{1}{b-a} \qquad a < x < b$$
$$= 0 \qquad \text{otherwise.}$$

한양대학교 HANYANG UNIVERSITY

Probability Distributions

- Practice
  - Draw 100 random sample from uniform (0,1)

In [9]:
```
from numpy import random

np.random.seed(10)

out1 = random.random(10000)
plt.hist(out1)
```



한양대학교 HANYANG UNIVERSITY

Probability Distributions

- Normal distribution

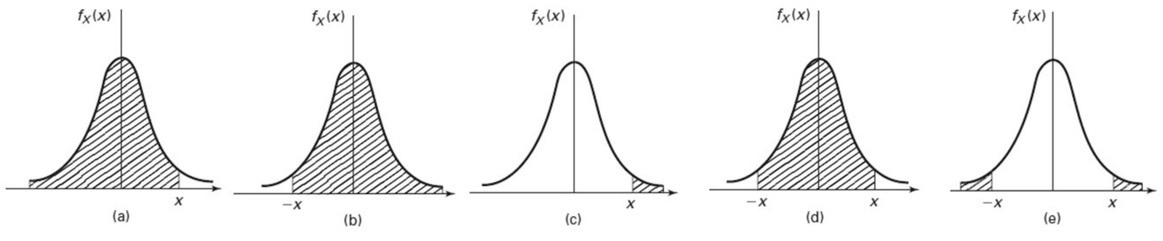
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}, -\infty < x < +\infty.$$

- Standard Normal distribution

$$f(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2}} \qquad , -\infty < x < \infty$$

Probability Distributions

- Normal distribution

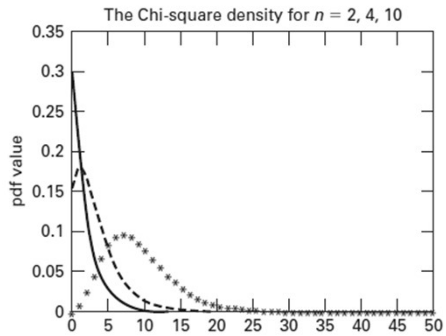


(a) $P[X \leq x]$; (b) $P[X > -x]$; (c) $P[X > x]$; (d) $P[-x < X \leq x]$; (e) $P[|X| > x]$.

Probability Distributions

- Chi-squre distribution

$$f_X(x) = K_\chi x^{\left(\frac{n}{2}\right)-1} e^{-\frac{x}{2}} u(x) \qquad K_\chi = \frac{1}{2^{n/2}\Gamma(n/2)}$$



한양대학교 HANYANG UNIVERSITY

Probability Distributions

- t-distribution

$$f(x) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\,\Gamma(\frac{v}{2})}\left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

한양대학교 HANYANG UNIVERSITY

Sampling Distributions

np.random.normal(*loc, scale, size*)
-. *loc* : mean
-. scale: standard deviation
-. *size* : output shape
-. Generate a random number from N(loc, scale)

np.random.chisquare(*df*, size)
-. *df* : degree of freedom
-. *size*: output shape

np.random.student_t(*df*, size)

Sampling Distributions

```
In [15]: from scipy.stats import binom, uniform, chi2, t, norm
```

-. *binom* : binomial distribution
-. *uniform*: uniform distribution
-. *chi2* : chi-square distribution
-. *t* : t-distribution
-. *norm*: normal distribution

Sampling Distributions

.rvs (*size*)
-. *Random sample*

.pdf (*x*) or .pmf (*x*)
-. *Probability density or probability mass*

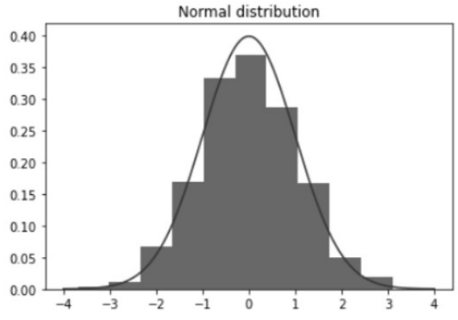
.cdf(*x*)
-. *Cumulative distribution*

.ppf(*prob*)
-. *point percentile*

Probability Distributions

- Practice (Normal distribution)

```
In [32]: rs = norm.rvs(loc=0, scale=1, size=1000)
         x = np.linspace(-4, 4, 1000)

         plt.hist(rs,  density=True, histtype='stepfilled')
         plt.plot(x, norm.pdf(x, loc=0, scale=1), 'r-')
         plt.title('Normal distribution')
         plt.show()
```



Probability Distributions

- Practice (Normal distribution)
  - The probability between 0 and 1 of standard normal distribution

```
In [34]: a = norm.cdf(0, loc=0, scale=1)
         b = norm.cdf(1, loc=0, scale = 1)
         print("The probability between 0 and 1 :", b-a)

         The probability between 0 and 1 : 0.3413447460685429
```

Probability Distributions

- Practice (Normal distribution)
  - Where the cumulative probability is 0.5:

```
In [35]: print("The point upto 0.5 :", norm.ppf(0.5, loc=0, scale=1))

         The point upto 0.5 : 0.0
```

Probability Distributions

- Central Limit Theorem

If $\bar{X}$ is the random sample $X_1, \dots, X_n$ of a size n from a distribution with a finite mean of μ and a finite positive variance σ^2,
Then, the distribution of

$$W = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

in the limit as $n \rightarrow \infty$

한양대학교 HANYANG UNIVERSITY