# Markov and Chebyshev’s inequality, Central Limit Theorem and Conditional probability and conditional expectation

# What do you think?

I toss a coin 1000 times. The probability that I get 14 consecutive heads is

| A | B | C |
|---|---|---|
| $< 10\%$ | $\approx 50\%$ | $> 90\%$ |

# Consecutive heads

Let $N$ be the number of occurrences of 14 consecutive heads in 1000-coin flips.

$$N = I_1 + \ldots + I_{987}$$

where $I_i$ is an indicator r.v. for the event

"14 consecutive heads starting at position $i$"

$$E[I_i] = P(I_i = 1) = 1/2^{14}$$

$$E[N] = 987 \cdot 1/2^{14} = 987/16384 \approx 0.0602$$

# Limit Theorems

- **Markov's Inequality:** *If X is a random variable that takes only nonnegative values, then for any value a > 0*

$$P(X \geq a) \leq \frac{E[X]}{a}$$

$E[N] \approx 0.0602$

$P[N \geq 1] \leq E[N] \,/\, 1 \leq 6\%.$

# Proof of Markov's inequality

For every non-negative random variable $X$: and every value $a$:

$$P(X \geq a) \leq E[X] \,/\, a.$$

$$E(X) = \int_0^\infty xp(x)dx = \int_0^a xp(x)dx + \int_a^\infty xp(x)dx$$

$$\geq \int_a^\infty ap(x)dx = a\int_a^\infty p(x)dx = aP(X \geq a)$$

$$P(X \geq a) \leq \frac{E(X)}{a}$$

# Example

1000 people throw their hats in the air. What is the probability at least 100 people get their hat back?

## Solution

$$N = I_1 + \ldots + I_{1000}$$

where $I_i$ is the indicator for the event that person $i$ gets their hat.
Then $E[I_i] = P(I_i = 1) = 1/n$

$$E[N] = n\,1/n = 1$$

$$P[N \geq 100] \leq E[N] \,/\, 100 = 1\%.$$

# What do you think?

- Suppose we know that the number of items produced in a factory during a week is a random variable with mean 500.
- If the variance of a week's production is known to equal 100, then what can be said about the probability that this week's production will be between 400 and 600?

# Chebyshev's Inequality

- What is the probability that the value of X is far from its expectation?
- Let X be a random variable with $E[X] = \mu, Var(X) = \sigma^2$

$$P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}, for\ any\ value\ k > 0$$

- Proof
  - Since $(X - \mu)^2$ is non-negative random variable, apply Markov's Inequality with $a = k^2$

$$P\left((x-\mu)^2 \geq k^2\right) \leq \frac{E[(x-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

  - Note that: $(X - \mu)^2 \geq k^2 \Leftrightarrow |X - \mu| \geq k$, yielding:

$$P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Markov's and Chebyshev's inequalities enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known.

# What do you think?

- Suppose we know that the number of items produced in a factory during a week is a random variable with mean 500.
- If the variance of a week's production is known to equal 100, then what can be said about the probability that this week's production will be between 400 and 600?
- Solution

$$P\{|X-500| \geq 100\} \leq \frac{\sigma^2}{(100)^2} = \frac{1}{100}$$

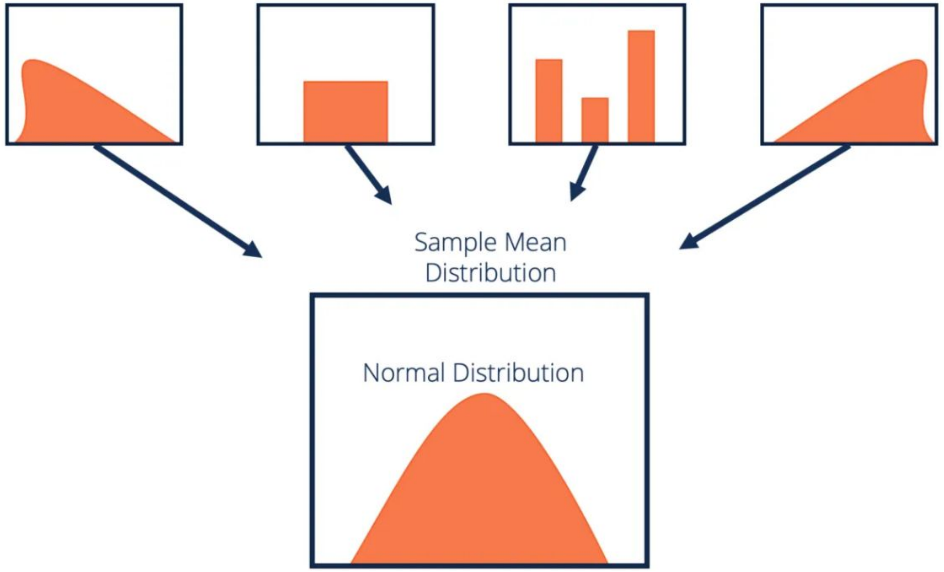
and so, the probability that this week's production will be between 400 and 600 is at least 0.99

# Central Limit Theorem

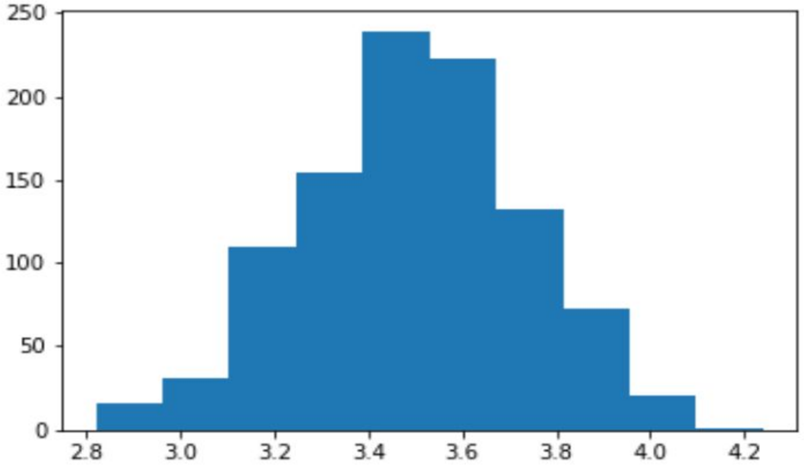
- **Recall**
  - **Observation**
    - **Result from one trial of an experiment.**
  - **Sample**
    - **Group of results gathered from separate independent trials.**
  - **Population**
    - **Space of all possible observations that could be seen from a trial.**
- **If we calculate the mean of a sample, it will be an estimate of the mean of the population distribution.**
- **But, like any estimate, it will be wrong and will contain some error.**
- **If we draw multiple independent samples, and calculate their means, the distribution of those means will form a Gaussian distribution.**

# Central Limit Theorem

- **The central limit theorem states that if you have a population with mean $\mu$ and standard deviation $\sigma$ and take sufficiently large random samples from the population with replacement, *then the distribution of the sample means will be approximately normally distributed.***




```python
# demonstration of the central limit theorem
from numpy.random import seed
from numpy.random import randint
from numpy import mean
from matplotlib import pyplot
# seed the random number generator
seed(1)
# calculate the mean of 50 dice rolls 1000 times
means = [mean(randint(1, 7, 50)) for _ in range(1000)]
# plot the distribution of sample means
pyplot.hist(means)
pyplot.show()
```



# Central Limit Theorem

- Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables, each with mean $\mu$ and variance $\sigma^2$. Then the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

- tends to the standard normal as $n \to \infty$. That is

$$P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx$$

# Conditional probability and conditional expectation

# Recall: The Discrete Case

- For any two events $E$ and $F$, the conditional probability of $E$ given $F$ is defined, as long as $P(F) > 0$, by

$$P(E|F) = \frac{P(EF)}{P(F)}$$

- Example:
  - Flip three different fair coins, then $S$={$HHH$, $HHT$, $HTH, HTT, THH, THT, TTH, TTT$}
  - Question: What is the probability that the first coin comes up heads when exactly two of the three coins came up heads.?

# The Discrete Case

- If X and Y are discrete random variables, then the **conditional probability mass function** of $X$ given that $Y = y$, by

$$p_{X|Y}(x|y) = P\{X = x|Y = y\}$$

$$= \frac{P\{X = x, Y = y\}}{P\{Y = y\}}$$

$$= \frac{p\{x, y\}}{p_Y\{y\}}, P\{Y = y\} > 0$$

# The Discrete Case

- Similarly, the **conditional probability distribution function** of $X$ given that $Y = y$ is defined, for all $y$ such that $P\{Y = y\} > 0$, by

$$F_{X|Y}(x|y) = P\{X \leq x|Y = y\}$$

$$= \sum_{a \leq x} p_{X|Y}(a|y)$$

# The Discrete Case

- The conditional expectation of X given that Y = y is defined by

$$E[X|Y=y] = \sum_x xP(X=x|Y=y)$$

$$= \sum_x x p_{X|Y}(x|y)$$

- if X is independent of Y, then
  $p_{X|Y}(x|y) = P(X=x|Y=y) = P(X=x)$

# Example

- **Suppose that $p(x,y)$, the joint probability mass function of X and Y, is given by $p(1,1) = 0.5, p(1,2) = 0.1, p(2,1) = 0.1, p(2,2) = 0.3$**
- **Question: Calculate the conditional probability mass function of X given that Y = 1**
- **Solution:**

$$p_Y(1) = \sum_x p(x,1) = p(1,1) + p(2,1) = 0.6$$

Hence,

$$\begin{aligned} p_{X|Y}(1|1) &= P\{X = 1|Y = 1\} \\ &= \frac{P\{X = 1, Y = 1\}}{P\{Y = 1\}} \\ &= \frac{p(1,1)}{p_Y(1)} \\ &= \frac{5}{6} \end{aligned}$$

Similarly,

$$p_{X|Y}(2|1) = \frac{p(2,1)}{p_Y(1)} = \frac{1}{6}$$

# Example

- If $X_1$ and $X_2$ are independent binomial random variables with respective parameters $(n_1, p)$ and $(n_2, p)$, calculate the conditional probability mass function of $X_1$ given that $X_1 + X_2 = m$.

$$P\{X_1 = k | X_1 + X_2 = m\} = \frac{P\{X_1 = k, X_1 + X_2 = m\}}{P\{X_1 + X_2 = m\}}$$

$$= \frac{P\{X_1 = k, X_2 = m - k\}}{P\{X_1 + X_2 = m\}}$$

$$= \frac{P\{X_1 = k\} P\{X_2 = m - k\}}{P\{X_1 + X_2 = m\}}$$

$$= \frac{\binom{n_1}{k} p^k q^{n_1 - k} \binom{n_2}{m-k} p^{m-k} q^{n_2 - m + k}}{\binom{n_1 + n_2}{m} p^m q^{n_1 + n_2 - m}}$$

$$P\{X_1 = k | X_1 + X_2 = m\} = \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1 + n_2}{m}}$$

# The Continuous Case

- If $X$ and $Y$ have a joint probability density function $f(x, y)$ , then the conditional probability density function of $X$, given that $Y = y$, is defined for all values of $y$ such that $f_Y(y) > 0$, by

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

# The Continuous Case

- $$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$
- The conditional expectation of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

$$E[X|Y = y\} = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

# Example

- **Suppose the joint density of X and Y is given by**

$$f(x,y) = f(x) = \begin{cases} 6xy(2-x-y), & 0<x<1, 0<y<1 \\ 0, & otherwise \end{cases}$$

- **Question: Compute the conditional expectation of X given that $Y = y$, where $0 < y < 1$.**

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{6xy(2-x-y)}{\int_0^1 6xy(2-x-y)\,dx} \\ &= \frac{6xy(2-x-y)}{y(4-3y)} \\ &= \frac{6x(2-x-y)}{4-3y} \end{aligned}$$

Hence,

$$\begin{aligned} E[X|Y=y] &= \int_0^1 \frac{6x^2(2-x-y)\,dx}{4-3y} \\ &= \frac{(2-y)2-\frac{6}{4}}{4-3y} \\ &= \frac{5-4y}{8-6y} \end{aligned}$$

# Reference


- Lecture notes on Probability Theory by Phanuel Mariano
- **Introduction to Probability Models, Sheldon M. Ross, Tenth Edition**