Topics:

- Variance and standard deviation.
- Markov inequality.
- Chebyshev inequality.
- Markov chains
Discrete random variables.

- **Example.** Let $X$ be a Binomial random variable with parameters $n = 200$ and $p = 0.035$. Find probabilities $P(X = 4)$ and $P(X = 6)$.

Here $P(X = 4) = C(200, 4)(0.035)^4(0.965)^{196} = 0.09003862196\ldots$ and $P(X = 6) = C(200, 6)(0.035)^6(0.965)^{194} = 0.1508966957\ldots$

- **Example.** Let $X$ be a Poisson random variable with parameter $\lambda = 7$. Find probabilities $P(X = 4)$ and $P(X = 6)$.

Here $P(X = 4) = e^{-7} \cdot \frac{7^4}{4!} = 0.09122619167\ldots$ and $P(X = 6) = e^{-7} \cdot \frac{7^6}{6!} = 0.1490027797\ldots$

- **Example.** Let $X$ be a geometric random variable with parameter $p = \frac{1}{7}$. Find probabilities $P(X = 4)$ and $P(X = 6)$.

Here $P(X = 4) = \frac{1}{7} \cdot \frac{6^3}{7^3} = 0.08996251562\ldots$ and $P(X = 6) = \frac{1}{7} \cdot \frac{6^5}{7^5} = \frac{7776}{117649} = 0.06609490943\ldots$
Variance and standard deviation.

• **Theorem.** Let \( X \) be a discrete random variable characterized by its probability mass function \( p(x) \). Then, for any real valued function \( g \), \( g(X) \) will also be a random variable, and

\[
E[g(X)] = \sum_{a: p(a)>0} g(a) \cdot p(a)
\]

• **Example.** We roll a fair die once, and square the outcome. Let \( X \) be a random variable representing the outcome. Then \( Y = X^2 \) will be a random variable representing the square of the outcome. Here

\[
p_X(1) = p_X(2) = p_X(3) = p_X(4) = p_X(5) = p_X(6) = \frac{1}{6}
\]

will be the probability mass function for \( X \), and

\[
p_Y(1) = p_Y(4) = p_Y(9) = p_Y(16) = p_Y(25) = p_Y(36) = \frac{1}{6}
\]

will be the probability mass function for \( Y \). Then

\[
E[Y] = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}
\]
**Variance and standard deviation.**

\[ E[g(X)] = \sum_{a: \ p(a) > 0} g(a) \cdot p(a) \]

- **Example.** We roll a fair die once, and square the outcome. Let \( X \) be a random variable representing the outcome. Then \( Y = X^2 \) will be a random variable representing the square of the outcome. Here

\[
p_X(1) = p_X(2) = p_X(3) = p_X(4) = p_X(5) = p_X(6) = \frac{1}{6}
\]

will be the probability mass function for \( X \), and

\[
p_Y(1) = p_Y(4) = p_Y(9) = p_Y(16) = p_Y(25) = p_Y(36) = \frac{1}{6}
\]

will be the probability mass function for \( Y \). Then

\[
E[Y] = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}
\]

Observe that

\[
\sum_{k=1}^{6} k^2 \cdot p_X(k) = \frac{91}{6}
\]

as well. Also observe that

\[
E[X^2] = \frac{91}{6} \neq (E[X])^2 = \left( \frac{7}{2} \right)^2 = \frac{49}{4}
\]
Variance and standard deviation.

- Given constants $\alpha$ and $\beta$,

$$E[\alpha X + \beta] = \alpha E[X] + \beta$$

**Proof:**

$$E[\alpha X + \beta] = \sum_{a: \ p(a) > 0} (\alpha a + \beta) \cdot p(a) = \alpha \sum_{a: \ p(a) > 0} a p(a) + \beta \sum_{a: \ p(a) > 0} p(a) = \alpha E[X] + \beta$$

Now, let $X$ be a random variable with mean $E[X] = \mu$.

- **Definition.** The variance of $X$ is

$$Var(X) = E[(X - \mu)^2]$$

Note that the variance is a mean square displacement from the mean $\mu$.

- **Definition.** The standard deviation of $X$ is

$$SD(X) = \sqrt{Var(X)} = \sqrt{E[(X - \mu)^2]}$$
Variance and standard deviation.

Let $X$ be a random variable with mean $E[X] = \mu$.

- **Definition.** The **standard deviation** of $X$ is

$$SD(X) = \sqrt{Var(X)} = \sqrt{E[(X - \mu)^2]}$$

Another notation: $\sigma(X)$ and $\sigma$.

- **Intuition:** $X = \mu \pm \sigma$

- **Example.** Let $X$ be a Binomial random variable with parameters $n$ and $p$. We know that $E[X] = np$. It will be shown that the variance

$$Var(X) = np(1 - p)$$

Thus

$$X = np \pm \sqrt{np(1 - p)}$$
Let $X$ be a Binomial random variable with $n = 100$ and $p = \frac{1}{2}$.

$$X = np \pm \sqrt{np(1-p)} = 50 \pm 5$$
Variance and standard deviation.

• Theorem. Let $X$ be a discrete random variable characterized by its probability mass function $p(x)$. Then, for any real valued function $g$, $g(X)$ will also be a random variable, and

$$E[g(X)] = \sum_{a: \ p(a)>0} g(a) \cdot p(a)$$

• Given constants $\alpha$ and $\beta$,

$$E[\alpha X + \beta] = \alpha E[X] + \beta$$

Now, let $X$ be a random variable with mean $E[X] = \mu$.

• Definition. The variance of $X$ is

$$Var(X) = E[(X - \mu)^2]$$

• Definition. The standard deviation of $X$ is

$$SD(X) = \sqrt{Var(X)} = \sqrt{E[(X - \mu)^2]}$$

• Theorem. The variance of $X$ is

$$Var(X) = E[X^2] - \mu^2$$
Variance and standard deviation.

• **Theorem.** The **standard deviation** of $X$ is

$$Var(X) = E[X^2] - \mu^2$$

• **Example.** Let $X$ be a Binomial random variable with parameters $n$ and $p$. Then

$$Var(X) = np(1 - p)$$

• **Example.** Let $X$ be a Poisson random variable with parameter $\lambda > 0$. Then

$$Var(X) = \lambda$$

• **Example.** Let $X$ be a geometric random variable with parameter $p$. Then

$$Var(X) = \frac{1 - p}{p^2}$$
**Markov inequality.**

- **Example.** When a certain lab experiment is performed, the outcome is an integer number on the scale from 0 to 20,000. Analyzing the outcomes of multiple identical experiments performed independently of each other it was noticed that the average value stays around 440. Suppose the threshold value is 10,000. If this is all we know, can we estimate how small is the probability that the outcome of one such experiment yields a value greater or equal to 10,000.

**Same stated in terms of random variables:** Let $X$ be a random variable, taking integer values from 0 to 20,000. We don’t know its probability mass function $p(k)$ ($k = 0, 1, 2, \ldots, 20K$). However we know that its expectation $E[X] = 440$. What can we say about the probability of going above the threshold

$$P(X \geq 10,000)$$

Can we bound it?
Markov inequality.

Same stated in terms of random variables: Let $X$ be a random variable, taking integer values from 0 to 20,000. We don’t know its probability mass function $p(k)$ ($k = 0, 1, 2, \ldots, 20K$). However we know that its expectation $E[X] = 440$. What can we say about the probability of going above the threshold

$P(X \geq 10,000)$

Can we bound it?

Theorem. (Markov inequality.) If $X$ is a random variable that takes only nonnegative values, then for any $\alpha > 0$,

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

Solution to the above example:

$P(X \geq 10,000) \leq \frac{440}{10,000} = 0.044$
**Markov inequality.**

**Theorem. (Markov inequality.)** If $X$ is a random variable that takes only nonnegative values, then for any $\alpha > 0$,

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

**Proof:**

$$P(X \geq \alpha) = \sum_{a: a \geq \alpha} p(a) \leq \sum_{a: a \geq \alpha} \frac{a}{\alpha} \cdot p(a) = \frac{1}{\alpha} \cdot \sum_{a: a \geq \alpha} a \cdot p(a) \leq \frac{1}{\alpha} \cdot \sum_{a: a \geq 0} a \cdot p(a) = \frac{E[X]}{\alpha}$$

**Example.** Let $X$ be a Binomial random variable with parameters $n = 2,500$ and $p = 0.2$. Use Markov inequality to give an upper bound on the following probability

$$P(X \geq 540) = \sum_{k=540}^{2,500} C(2500, k) \cdot (0.2)^k \cdot (0.8)^{2,500-k}$$
Markov inequality.

Theorem. (Markov inequality.) If $X$ is a random variable that takes only nonnegative values, then for any $\alpha > 0$,

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

- **Example.** Let $X$ be a Binomial random variable with parameters $n = 2,500$ and $p = 0.2$. Use Markov inequality to give an upper bound on the following probability

$$P(X \geq 540) = \sum_{k=540}^{2,500} C(2500, k) \cdot (0.2)^k \cdot (0.8)^{2,500-k}$$

**Solution:** Here $E[X] = np = 500$. Thus

$$P(X \geq 540) \leq \frac{500}{540} = 0.925\ldots$$

- **Comment:** Here we also know the standard deviation $\sigma = \sqrt{np(1-p)} = 20$. Thus we know that $X = \mu \pm \sigma = 500 \pm 20$, making us believe that $P(X \geq 540)$ is much smaller than 92.5%.
We know that $X = \mu \pm \sigma = 500 \pm 20$, making us believe that $P(X \geq 540)$ is much smaller than 92.5%.
In fact, $P(X \geq 540) \approx 0.0249 << 0.925$. 
Chebyshev inequality.

**Theorem. (Chebyshev inequality.)** If $X$ is a random variable with finite mean $\mu$ and variance, then for any $\kappa > 0$,

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

**Example.** Let $X$ be a Binomial random variable with parameters $n = 2,500$ and $p = 0.2$. Give an upper bound on the following probability

$$P(X \geq 540) = \sum_{k=540}^{2,500} C(2500, k) \cdot (0.2)^k \cdot (0.8)^{2,500-k}$$

**Solution:** Here $\mu = np = 500$ and $Var(X) = np(1-p) = 400$. Thus

$$P(X \geq 540) = P(X - \mu \geq 40) \leq P(|X - \mu| \geq 40) \leq \frac{400}{40^2} = 0.25$$
Markov and Chebyshev inequalities.

**Theorem. (Markov inequality.)** If $X$ is a random variable that takes only nonnegative values, then for any $\alpha > 0$,

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

**Theorem. (Chebyshev inequality.)** If $X$ is a random variable with finite mean $\mu$ and variance, then for any $\kappa > 0$,

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

**Proof:** Let $Y = (X - \mu)^2$, then $E[Y] = Var(X)$ and

$$P(|X - \mu| \geq \kappa) = P((X - \mu)^2 \geq \kappa^2) = P(Y \geq \kappa^2) \leq \frac{E[Y]}{\kappa^2} = \frac{Var(X)}{\kappa^2}$$

using Markov inequality for $Y$, since $Y$ is a nonnegative random variable.
Markov chains.

Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ with values in the discrete state space $S$.

The sequence $\{X_t\}_{t=0,1,\ldots}$ is said to be a discrete time Markov chain if it satisfies the following property, known as Markov property:

$$P(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{t+1} = j \mid X_t = i)$$

A Markov chain $\{X_t\}_{t=0,1,\ldots}$ is said to be time homogeneous if

$$P(X_{t+1} = j \mid X_t = i) = p(i,j) \quad \text{for all } t = 0, 1, 2, \ldots$$
Wright-Fisher Model.

Model: Let $X_t$ denote the number of alleles $A$ in $t$-th generation. Thus, there are $2N - X_t$ of alleles $a$.

For each of the $2N$ copies of the locus in $(t+1)$-st generation, one of the $2N$ alleles in $t$-th generation is selected uniformly at random. Thus, in these $2N$ Bernoulli trials, allele $A$ comes up with probability

$$p = \frac{X_t}{2N}$$

and allele $a$ comes up with probability $1 - p$.

Hence, given that we know $X_t$, variable $X_{t+1}$ representing the number of alleles $A$ in $(t+1)$-st generation is a Bernoulli random variables with $2N$ trials and $p = \frac{X_t}{2N}$, i.e.

$$P(X_{t+1} = k \mid X_t = m) = C(2N, k)p^k(1 - p)^{2N-k}, \quad \text{where } p = \frac{m}{2N}.$$
Wright-Fisher Model.

Hence, given that we know $X_t$, variable $X_{t+1}$ representing the number of alleles $A$ in $(t+1)$-st generation is a Bernoulli random variables with $2N$ trials and $p = \frac{X_t}{2N}$, i.e.

$$P(X_{t+1} = k \mid X_t = m) = \binom{2N}{k} p^k (1 - p)^{2N-k}, \text{ where } p = \frac{m}{2N}.$$ 

Urn model and fixation time: