MTH 463/563

Lectures 11 - 16

Yevgeniy Kovchegov
Oregon State University

Topics:

- Introduction to random variables.
- Binomial random variables.
- Expectation of a discrete random variable.
- Poisson random variables.
- Poisson vs Binomial.
- Geometric random variables.

- Examples with discrete random variables.
- Variance and standard deviation.
- Markov inequality.
- Chebyshev inequality.
- Review and examples.

Introduction to random variables.

Consider a sample space $\mathcal{S}$ and a probability function P.

- **Definition.** A function from $\mathcal{S}$ to $\mathbb{R}$ is a **random variable**.

- **Example.** Roll two fair dice. Let $X(i,j) = i + j$ for each outcome (i,j) in $\mathcal{S}$. Then X is a random variable representing the sum of the digits on the dice.

X :

1,1	1,2	1,3	1,4	1,5	1, 6
2,1	2,2	2,3	2,4	2, 5	2,6
3,1	3,2	3,3	3, 4	3,5	3,6
4,1	4,2	4, 3	4,4	4,5	4,6
5,1	5, 2	5,3	5,4	5,5	5,6
6, 1	6,2	6,3	6,4	6,5	6,6

$$\longrightarrow \left\{ \begin{matrix} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{matrix} \right\}$$

Introduction to random variables.

• **Example.** Roll two fair dice. Let $X(i,j) = i + j$ for each outcome (i,j) in $\mathcal{S}$. Then X is a random variable representing the sum of the digits on the dice.

X :

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

$$\longrightarrow \left\{ \begin{matrix} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{matrix} \right\}$$

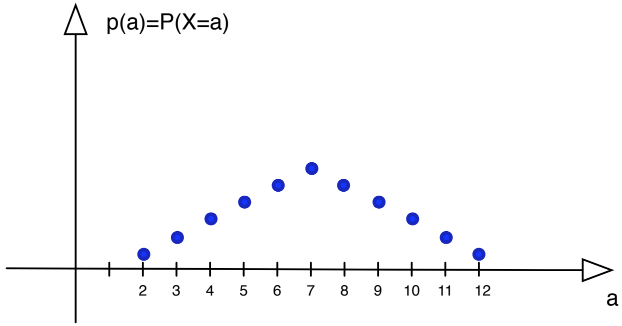
Here, for example, $X(3,1) = 4$ and $X(5,6) = 11$.

We are interested in finding the following probabilities:

$$p(a) = P(X = a) \quad \text{for} \quad a = 2, 3, \ldots, 12$$

Introduction to random variables.

- **Example.** Roll two fair dice. Let $X(i,j) = i + j$ for each outcome (i,j) in $\mathcal{S}$. Then X is a random variable representing the sum of the digits on the dice.



We are interested in finding the following probabilities:

$$p(a) = P(X = a) \quad \text{for} \quad a = 2, 3, \ldots, 12$$

$$p(2) = \tfrac{1}{36},\ p(3) = \tfrac{2}{36},\ p(4) = \tfrac{3}{36},\ p(5) = \tfrac{4}{36},\ p(6) = \tfrac{5}{36},\ p(7) = \tfrac{6}{36}$$

$$p(8) = \tfrac{5}{36},\ p(9) = \tfrac{4}{36},\ p(10) = \tfrac{3}{36},\ p(11) = \tfrac{2}{36},\ p(12) = \tfrac{1}{36}$$

Introduction to random variables.

Let X a discrete random variable. That is X assumes a discrete (countable) number of values.

- **Definition.** Function $p(a) = P(X = a)$ is called the probability mass function (or distribution function).

- **Definition.** Function $F(a) = P(X \leq a)$ is called the cumulative distribution function.

- **Note.** $$\sum_{a:\ p(a)>0} p(a) = 1$$

  In the previous example, $p(2)+p(3)+\cdots+p(12) = 1$.

- **Note.** $0 \leq F(a) \leq 1$

- **Note.** $F(a) = \sum_{x:\ x \leq a} p(x)$

Bernoulli trials and Bernoulli random variables.

For a given $0 \leq p \leq 1$, a Bernoulli trial is an experiment with exactly two possible outcomes, success and failure, in which the probability of success is p and probability of failure is $1 - p$.

Here, the sample space $\mathcal{S}$ consists of the two outcomes, success and failure, and

$$P(\text{success}) = p \quad \text{and} \quad P(\text{failure}) = 1 - p$$

Bernoulli random variable X with parameter p counts the number of successes after one Bernoulli trial, and thus,

$$P(X = 1) = p \quad \text{and} \quad P(X = 0) = 1 - p$$

Expectation of a discrete random variable.

• **Definition.** Let X be a discrete random variable with the probability mass function $p(x)$. Then its expected value is

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

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

$$p(1) = P(X = 1) = p \quad \text{and} \quad p(0) = P(X = 0) = 1 - p$$

and

$$E[X] = 0 \cdot p(0) + 1 \cdot p(1) = p$$

• **Example.** Roll two fair dice. Let X represent the sum of the digits on the dice. Then

$$E[X] = 2 \cdot p(2) + 3 \cdot p(3) + \cdots + 12 \cdot p(12) = \frac{252}{36} = 7$$

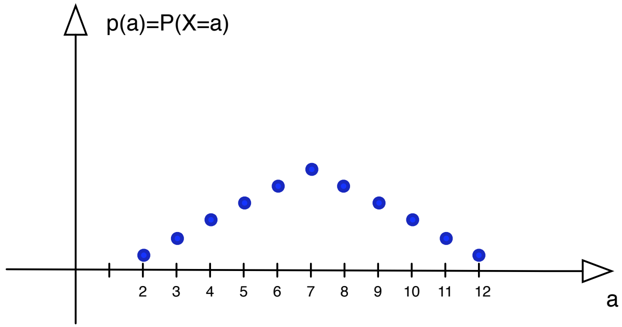
Expectation of a discrete random variable.

• **Definition.** Let X be a discrete random variable with the probability mass function $p(x)$. Then its **expected value** is

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

• **Example.** Roll two fair dice. Let X represent the sum of the digits on the dice. Then

$$E[X] = 2 \cdot p(2) + 3 \cdot p(3) + \cdots + 12 \cdot p(12) = \frac{252}{36} = 7$$



This corresponds to a **center of mass** of $p(a)$.

Binomial random variable. Recall the following example.

- **Example.** Consider performing independent Bernoulli trials, each with probability p of success and probability $1-p$ of failure. Let X be a random variable representing the number of successes in n Bernoulli trials. Find $P(X=k)$ for $k=0,1,\ldots,n$.

- **Solution.**

Each outcome with k successes and $n-k$ failures, its probability

$$P(\underbrace{SFSS\ldots FFS}_{k\ S's\ \text{and}\ n-k\ F's}) = p^k(1-p)^{n-k}$$

and $$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for each } k=0,1,\ldots,n$$

because there are $\binom{n}{k}$ such outcomes.

- **Definition.** The random variable X in the above example is the binomial random variable with parameters (n,p).

Check: $$\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} = \big(p+(1-p)\big)^n = 1^n = 1.$$

Expectation of a discrete random variable.

Let X be a binomial random variable with parameters (n, p). Then its probability mass function is known to be

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for each } k = 0, 1, \dots, n$$

• **Definition.** Let X be a discrete random variable with the probability mass function $p(x)$. Then its expected value is

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

• **Example.** Let X be a binomial random variable with parameters (n, p). Then

$$E[X] = \sum_{k=0}^{n} k \cdot p(k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = ?$$

Expectation of a discrete random variable.

• **Example.** Let X be a **binomial random variable** with parameters (n, p). Then $E[X] = np$ since

$$E[X] = \sum_{k=0}^{n} k \cdot p(k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} \, p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} \, p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \, p^k (1-p)^{n-k}$$

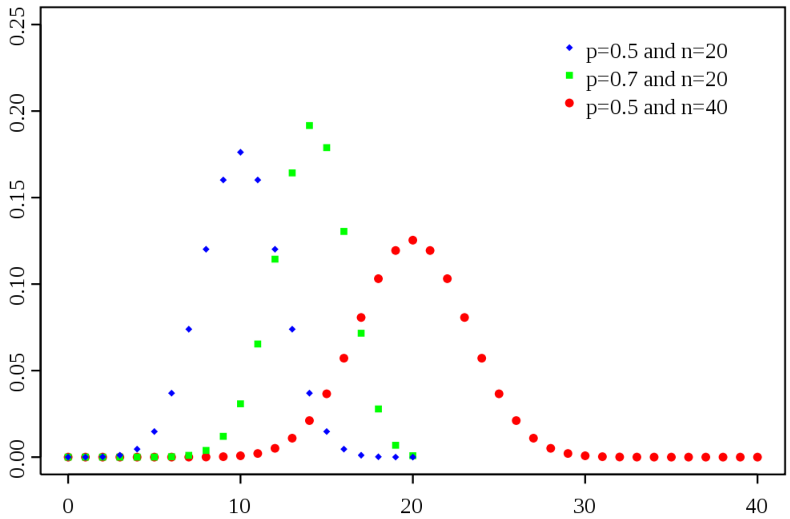
$$= \sum_{j=0}^{n-1} \frac{n!}{j!(n-1-j)!} \, p^{j+1} (1-p)^{n-1-j} = np \cdot \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} \, p^j (1-p)^{n-1-j},$$

where the new index $j = k - 1$. Thus

$$E[X] = np \cdot \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} \, p^j (1-p)^{n-1-j} = np \cdot \sum_{j=0}^{n-1} \binom{n-1}{j} \, p^j (1-p)^{n-1-j}$$

$$= np \cdot (p + (1-p))^{n-1} = np \qquad \text{by the Binomial theorem}$$

Binomial random variable.



Picture credit: Wikipedia.org

$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ for each $k = 0, 1, \ldots, n$ and $E[X] = np$

Poisson random variable.

- **Recall that** $\sum_{k=0}^{\infty} \frac{a^k}{k!} = e^a$ for all $-\infty < a < +\infty$

- **Definition.** Let $\lambda > 0$. A discrete random variable such that its probability mass function

$$p(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad \text{for each } k = 0, 1, \ldots$$

is a **Poisson random variable** with parameter $\lambda > 0$.

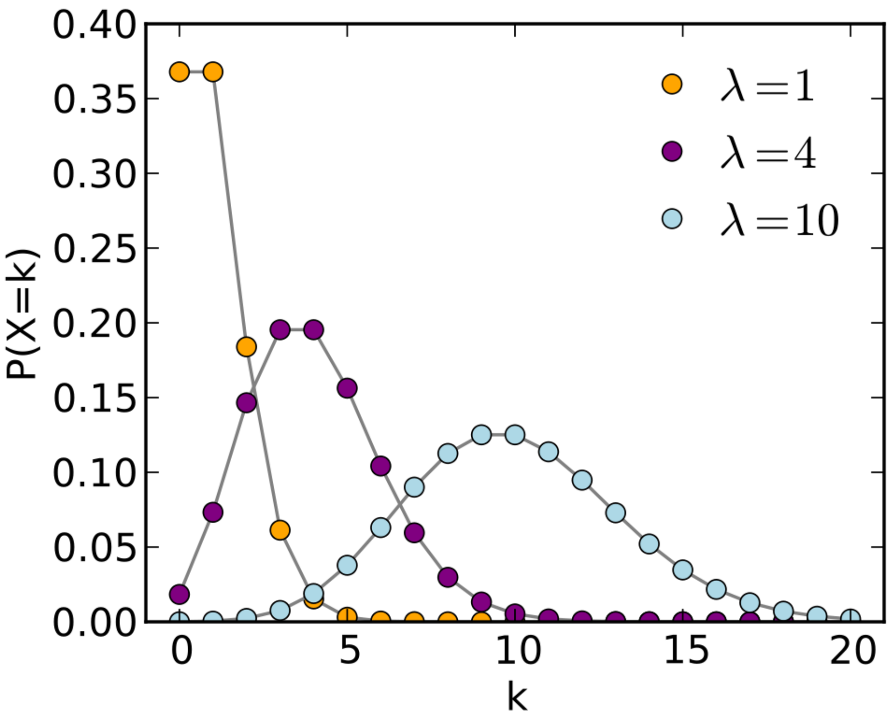
- Function $p(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ is **Poisson distribution**

- Check $\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$

- **Expectation:** Let X be a Poisson random variable with parameter λ. Then $E[X] = \lambda$ since

$$E[X] = \sum_{k=0}^{\infty} k \cdot p(k) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

Poisson random variable.



Picture credit: Wikipedia.org

$p(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ for each $k = 0, 1, \ldots$ and $E[X] = \lambda$

Poisson vs Binomial.

Let $\lambda > 0$ be given. Suppose Y is a Poisson random variable with parameter λ. Then its probability mass function

$$P(Y = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad \text{for each } k = 0, 1, \ldots$$

Now, let S_n be a Binomial random variable with parameters n and $p = \frac{\lambda}{n}$. Then its probability mass function

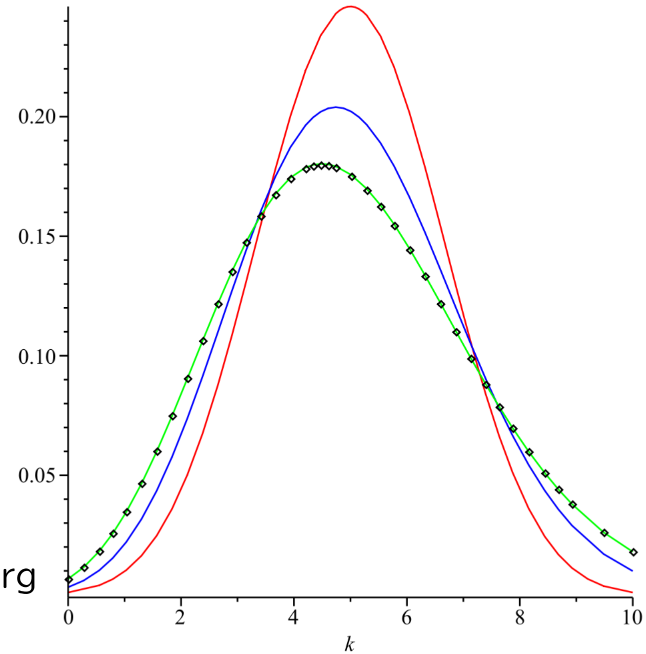
$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad \text{for each } k = 0, 1, \ldots, n$$

- **Theorem.** Consider integer $k \geq 0$. Then for n large enough,

$$P(S_n = k) \approx P(Y = k)$$

Namely, $\lim_{n \to \infty} P(S_n = k) = P(Y = k)$

Poisson vs Binomial.



Picture credit: Wikipedia.org

Dots: Poisson($\lambda = 5$) Red: Binomial($n = 10,\ p = \frac{1}{2}$)

Blue: Binomial($n = 20,\ p = \frac{1}{4}$)

Green: Binomial($n = 1000,\ p = \frac{1}{200}$)

Poisson vs Binomial.

Let $\lambda > 0$ be given. Suppose Y is a Poisson random variable with parameter λ and S_n is a Binomial random variable with parameters n and $p = \frac{\lambda}{n}$.

• **Theorem.** For a given integer $k \geq 0$, $\lim_{n\to\infty} P(S_n = k) = P(Y = k)$.

Thus, for n large enough, $P(S_n = k) \approx P(Y = k)$.

Proof: k is fixed, and

$$P(S_n = k) = \frac{n!}{k!(n-k)!} \cdot \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} \cdot \frac{n!}{(n-k)!} \cdot \frac{1}{n^k} \cdot \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^k}$$

$$= \frac{\lambda^k}{k!} \cdot \frac{(n-k+1)(n-k+2)\dots n}{n^k} \cdot \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^k} \longrightarrow e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad \text{as } n \to \infty$$

since $\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$, $\frac{(n-k+1)(n-k+2)\dots n}{n^k} \to 1$, and $\left(1 - \frac{\lambda}{n}\right)^k \to 1^k = 1$.

Geometric random variables.

- **Example.** Consider performing independent Bernoulli trials, each with probability p of success and probability $1-p$ of failure. Let X be a random variable representing the number of trials until the first success. Find $P(X=k)$ for $k=1,2,\ldots$.

- **Solution.**

The sample space $\mathcal{S}$ consists of the outcomes of infinitely many Bernoulli trials. For example $FFSFSSFS\ldots$ is one such outcome. Here X is a function from the sample space $\mathcal{S}$ to $\mathbb{R}$, and here

$$X(FFSFSSFS\ldots)=3$$

$$P(X=3)=P(F_1F_2S_3)=P(F_1)\cdot P(F_2)\cdot P(S_3)=p\cdot(1-p)^2$$

and $$P(X=k)=P(F_1)\cdot\dots\cdot P(F_{k-1})\cdot P(S_k)=p\cdot(1-p)^{k-1}\quad\text{for each } k=1,2,\ldots$$

- **Definition.** The random variable X in the above example is called a geometric random variable with parameter p.

Geometric random variables.

A **geometric random variable** with parameter p is characterized by a probability mass function,

$$p(k) = p \cdot (1-p)^{k-1} \quad \text{for each } k = 1, 2, \ldots$$

- We need to check that $\sum_{k=1}^{\infty} p(k) = 1$.

Geometric series: $\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \ldots$

Claim: For $x \neq 1$, $\sum_{k=0}^{n} x^k = 1 + x + x^2 + x^3 + \cdots + x^n = \frac{1-x^{n+1}}{1-x}$

Proof: $(1-x)(1+x+x^2+x^3+\cdots+x^n) = [1+x+x^2+x^3+\cdots+x^n] - [x+x^2+x^3+\cdots+x^n+x^{n+1}] = 1-x^{n+1}$

Summing the geometric series: For $|x| < 1$,

$$\sum_{k=0}^{\infty} x^k = \lim_{n\to\infty} \left(\sum_{k=0}^{n} x^k \right) = \lim_{n\to\infty} \frac{1-x^{n+1}}{1-x} = \frac{1}{1-x}$$

Geometric random variables.

Summing the geometric series: For $|x| < 1$,

$$\sum_{k=0}^{\infty} x^k = \lim_{n\to\infty} \left(\sum_{k=0}^{n} x^k \right) = \lim_{n\to\infty} \frac{1 - x^{n+1}}{1 - x} = \frac{1}{1 - x}$$

A **geometric random variable** with parameter p is characterized by a probability mass function,

$$p(k) = p \cdot (1 - p)^{k-1} \quad \text{for each } k = 1, 2, \ldots$$

- We need to check that $\sum_{k=1}^{\infty} p(k) = 1$.

$$\sum_{k=1}^{\infty} p(k) = p \cdot \sum_{k=1}^{\infty} (1 - p)^{k-1} = p \cdot \sum_{j=0}^{\infty} (1 - p)^j = p \cdot \frac{1}{1 - (1 - p)} = 1,$$

where $j = k - 1$.

Geometric random variables.

A **geometric random variable** with parameter p is characterized by a probability mass function,

$$p(k) = p \cdot (1-p)^{k-1} \quad \text{for each } k = 1, 2, \ldots$$

We need to find its expectation $E[X] = \sum_{k=1}^{\infty} k \cdot p(k)$.

$$E[X] = \sum_{k=1}^{\infty} k \cdot p(k) = p \cdot \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} = ?$$

Geometric random variables.

A **geometric random variable** with parameter p is characterized by a probability mass function,

$$p(k) = p \cdot (1-p)^{k-1} \quad \text{for each } k = 1, 2, \ldots$$

We need to find its expectation $E[X] = \sum_{k=1}^{\infty} k \cdot p(k)$.

$$E[X] = \sum_{k=1}^{\infty} k \cdot p(k) = p \cdot \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} = ?$$

Here $\sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$ for $|x| < 1$ as

$$\sum_{k=1}^{\infty} k x^{k-1} = \sum_{k=0}^{\infty} k x^{k-1} = \sum_{k=0}^{\infty} (x^k)' = \left(\sum_{k=0}^{\infty} x^k \right)' = \left(\frac{1}{1-x} \right)' = \frac{1}{(1-x)^2}$$

and therefore $E[X] = p \cdot \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} = p \cdot \frac{1}{(1-(1-p))^2} = \frac{1}{p}$

Geometric random variables.

Alternative proof of $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$ for $|x| < 1$.

$1 + 2x + 3x^2 + 4x^3 + 5x^4 + \cdots =$

$$\begin{array}{cccccccl}
1+ & x+ & x^2+ & x^3+ & x^4+ & \cdots & = & \frac{1}{1-x} \\
 & + & + & + & + & & & \\
 & x+ & x^2+ & x^3+ & x^4+ & \cdots & = & \frac{x}{1-x} \\
 & & + & + & + & & & \\
 & & x^2+ & x^3+ & x^4+ & \cdots & = & \frac{x^2}{1-x} \\
 & & & + & + & & & \\
 & & & x^3+ & x^4+ & \cdots & = & \frac{x^3}{1-x} \\
 & & & & + & & & \\
 & & & & x^4+ & \ldots & & \vdots
\end{array}$$

$$\frac{1}{1-x} + \frac{x}{1-x} + \frac{x^2}{1-x} + \frac{x^3}{1-x} + \cdots = \frac{1}{1-x} \cdot (1 + x + x^2 + \ldots) = \frac{1}{(1-x)^2}$$

Geometric random variables.

Let X be **geometric random variable** with parameter p. Then

$$p(k) = p \cdot (1-p)^{k-1} \quad \text{for each } k = 1, 2, \ldots$$

and

$$E[X] = \frac{1}{p}$$

- Prove the following *memorylessenss* property:

$$P(X = n + k \mid X > n) = P(X = k)$$

for any two positive integers n and k.

We observe that here, $P(X = n + k \mid X > n) = \frac{P(X=n+k)}{P(X>n)}$

- Apply the above to coin tossing. Give an example.

Geometric random variables.

Let X be **geometric random variable** with parameter p. Then

$$p(k) = p \cdot (1-p)^{k-1} \quad \text{for each } k = 1, 2, \ldots$$

and

$$E[X] = \frac{1}{p}$$

- **Example.** Find probability $P(X \geq 10)$.

- **Example.** Let $p = \frac{1}{2}$. Find probability $P(X \geq 20)$.

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

• **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

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

• **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.

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

$$E[g(X)] = \sum_{x:\, p_X(x)>0} g(x)\, p_X(x)$$

Proof: Let $Y = g(X)$. We find the probability mass function $p_Y(y)$ of Y:

$$p_Y(y) = P\big(g(X) = y\big) = \sum_{x:\, g(x)=y} P(X = x) = \sum_{x:\, g(x)=y} p_X(x)$$

as $\big\{g(X) = y\big\} = \bigcup_{x:\, g(x)=y} \{X = x\}$ is a union of disjoint events.

Thus, $$E[Y] = \sum_y y\, p_Y(y) = \sum_y \left(y \sum_{x:\, g(x)=y} p_X(x) \right) = \sum_y \left(\sum_{x:\, g(x)=y} y\, p_X(x) \right)$$

$$= \sum_y \left(\sum_{x:\, g(x)=y} g(x)\, p_X(x) \right) = \sum_x g(x)\, p_X(x)$$

Variance and standard deviation.

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

• **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

$$E[X^2] = \frac{91}{6} \quad \neq \quad \big(E[X]\big)^2 = \left(\frac{7}{2}\right)^2 = \frac{49}{4}$$

Examples.

• **Problem.** Random variable X has the following probability mass function

$$p_X(x) = \begin{cases} \frac{1}{8} & \text{if } x = -2 \\ \frac{5}{8} & \text{if } x = 2 \\ \frac{1}{4} & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

That is $p_X(-2) = \frac{1}{8}$, $p_X(2) = \frac{5}{8}$ and $p_X(3) = \frac{1}{4}$.

Compute $E[X]$ and $E[X^2]$. Hint: Recall that $E[g(X)] = \sum_{x:\, p_X(x)>0} g(x)\, p_X(x)$.

Solution: $E[X] = (-2) \cdot p_X(-2) + 2 \cdot p_X(2) + 3 \cdot p_X(3) = \frac{7}{4}$

$$E[X^2] = (-2)^2 \cdot p_X(-2) + 2^2 \cdot p_X(2) + 3^2 \cdot p_X(3) = \frac{21}{4}$$

Examples.

• **Example.** Let X be binomial random variable with parameters $n = 20$ and $p = \frac{1}{4}$. Use the binomial theorem to compute $E[2^X]$.

Solution:

$$E[2^X] = \sum_{k=0}^{n} 2^k \cdot p(k) = \sum_{k=0}^{n} 2^k \cdot \binom{n}{k} \cdot p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \cdot (2p)^k (1-p)^{n-k}$$

$$= (2p + (1-p))^n = (1+p)^n = \left(\frac{5}{4}\right)^{20} = 86.7361738$$

Variance and standard deviation.

- Given constants α and β,

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

Proof:

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

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

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

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

Note that the variance is a mean square displacement from the mean μ.

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

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

Variance and standard deviation.

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

• **Definition.** The **standard deviation** of X is

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

Another notation: $\sigma(X)$ and σ.

• **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

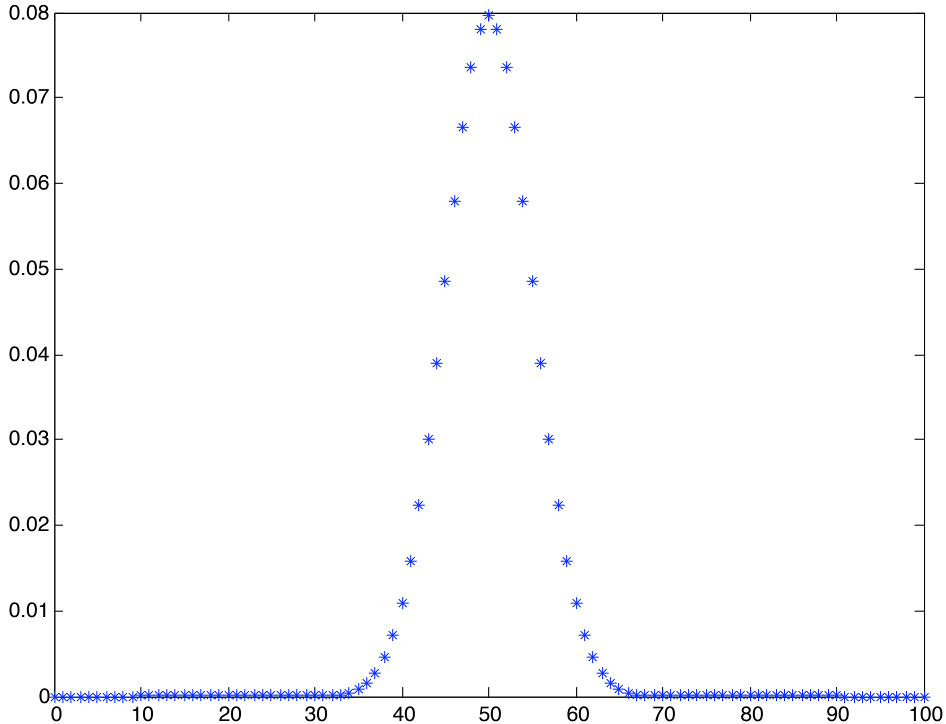
$$Var(X) = np(1-p)$$

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.

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

- **Theorem.** The **variance** of X equals

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

Proof:

$$Var(X) = E\Big[(X-\mu)^2\Big] = \sum_{a:\, p(a)>0} (a-\mu)^2 \cdot p(a) = \sum_{a:\, p(a)>0} (a^2 - 2\mu a + \mu^2) \cdot p(a)$$

$$= \sum_{a:\, p(a)>0} a^2 \cdot p(a) - 2\mu \cdot \sum_{a:\, p(a)>0} a \cdot p(a) + \mu^2 \cdot \sum_{a:\, p(a)>0} p(a)$$

$$= \sum_{a:\, p(a)>0} a^2 \cdot p(a) - 2\mu \cdot \mu + \mu^2 \cdot 1 = E[X^2] - 2\mu^2 + \mu^2$$

$$= E[X^2] - \mu^2$$

Variance and standard deviation.

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

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

Solution: Here $\mu = np$ and

$$Var(X) = E[X^2] - \mu^2 = \sum_{k=0}^{n} k^2 \cdot \binom{n}{k} p^k (1-p)^{n-k} - \mu^2$$

$$= \sum_{k=0}^{n} (k^2 - k) \cdot \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} - \mu^2$$

$$= \sum_{k=2}^{n} k(k-1) \cdot \binom{n}{k} p^k (1-p)^{n-k} + \mu - \mu^2 = \sum_{k=2}^{n} k(k-1) \cdot \frac{n!}{k!(n-k)!} \cdot p^k (1-p)^{n-k} + \mu - \mu^2$$

$$= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} \cdot p^k (1-p)^{n-k} + \mu - \mu^2$$

Variance and standard deviation.

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

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

Solution (continued): Here $\mu = np$ and

$$Var(X) = \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} \cdot p^k (1-p)^{n-k} + \mu - \mu^2$$

$$= p^2 \cdot n(n-1) \cdot \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} \cdot p^{k-2}(1-p)^{n-k} + \mu - \mu^2$$

$$= p^2 \cdot n(n-1) \cdot \sum_{j=0}^{n-2} \binom{n-2}{j} \cdot p^j (1-p)^{(n-2)-j} + \mu - \mu^2, \qquad \text{where } j = k-2$$

$$= p^2 \cdot n(n-1) \cdot (p + (1-p))^{n-2} + \mu - \mu^2 = p^2 \cdot n(n-1) + \mu - \mu^2$$

$$= p^2 \cdot (n^2 - n) + np - (np)^2 = -np^2 + np = np(1-p)$$

Variance and standard deviation.

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

$$Var(X) = \lambda$$

Solution: Here $\mu = \lambda$ and

$$Var(X) = E[X^2] - \mu^2 = \sum_{k=0}^{\infty} k^2 \cdot e^{-\lambda}\frac{\lambda^k}{k!} - \mu^2$$

$$= \sum_{k=0}^{\infty}(k^2 - k) \cdot e^{-\lambda}\frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} k \cdot e^{-\lambda}\frac{\lambda^k}{k!} - \mu^2$$

$$= \sum_{k=2}^{\infty} k(k-1) \cdot e^{-\lambda}\frac{\lambda^k}{k!} + \mu - \mu^2 = \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-2)!} + \mu - \mu^2$$

$$= \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^{j+2}}{j!} + \mu - \mu^2 = \lambda^2 \cdot e^{-\lambda} \cdot \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} + \mu - \mu^2, \qquad \text{where } j = k-2$$

$$= \lambda^2 \cdot e^{-\lambda} \cdot e^{\lambda} + \lambda - \lambda^2 = \lambda$$

Variance and standard deviation.

• **Example.** Let X be a geometric random variable with parameter p. Show that

$$Var(X) = \frac{1-p}{p^2}$$

Solution: Here $\mu = \frac{1}{p}$ and

$$Var(X) = E[X^2] - \mu^2 = \sum_{k=1}^{\infty} k^2 \cdot p \cdot (1-p)^{k-1} - \mu^2$$

$$= \sum_{k=1}^{\infty} k(k-1) \cdot p \cdot (1-p)^{k-1} + \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} - \mu^2$$

$$= p \cdot (1-p) \cdot \sum_{k=0}^{\infty} k(k-1) \cdot (1-p)^{k-2} + \mu - \mu^2$$

Variance and standard deviation.

• **Example.** Let X be a geometric random variable with parameter p. Show that

$$Var(X) = \frac{1-p}{p^2}$$

Solution (continued): Here $\mu = \frac{1}{p}$ and

$$Var(X) = p \cdot (1-p) \cdot \sum_{k=0}^{\infty} k(k-1) \cdot (1-p)^{k-2} + \mu - \mu^2$$

Now, for $|x| < 1$,

$$\sum_{k=0}^{\infty} k(k-1) \cdot x^{k-2} = \sum_{k=0}^{\infty} \left(x^k\right)'' = \frac{d^2}{dx^2}\left(\sum_{k=0}^{\infty} x^k\right) = \frac{d^2}{dx^2}\left(\frac{1}{1-x}\right) = \frac{2}{(1-x)^3}$$

Hence,

$$Var(X) = p \cdot (1-p) \cdot \frac{2}{p^3} + \mu - \mu^2 = 2 \cdot \frac{1-p}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Variance and standard deviation.

• **Theorem.** The **variance** of X equals

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

• **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

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

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

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

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:

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

• **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

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

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

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

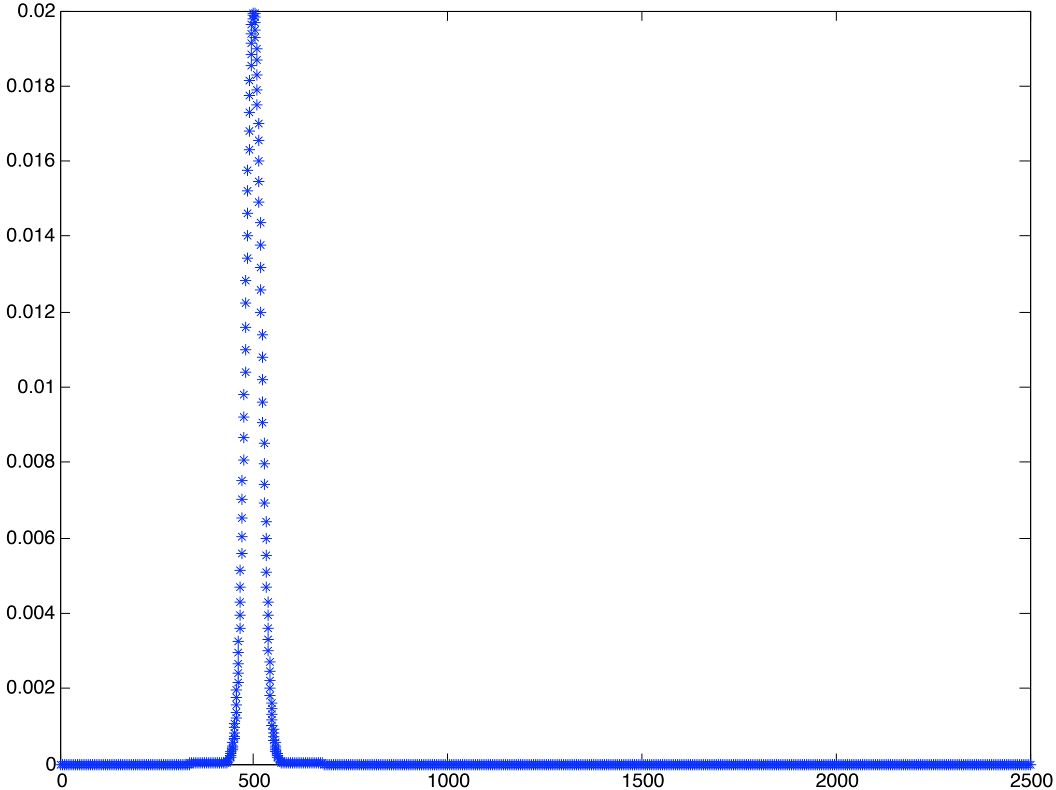
Solution: Here $E[X] = np = 500$. Thus

$$P(X \geq 540) \;\leq\; \frac{500}{540} = 0.\overline{925}\ldots$$

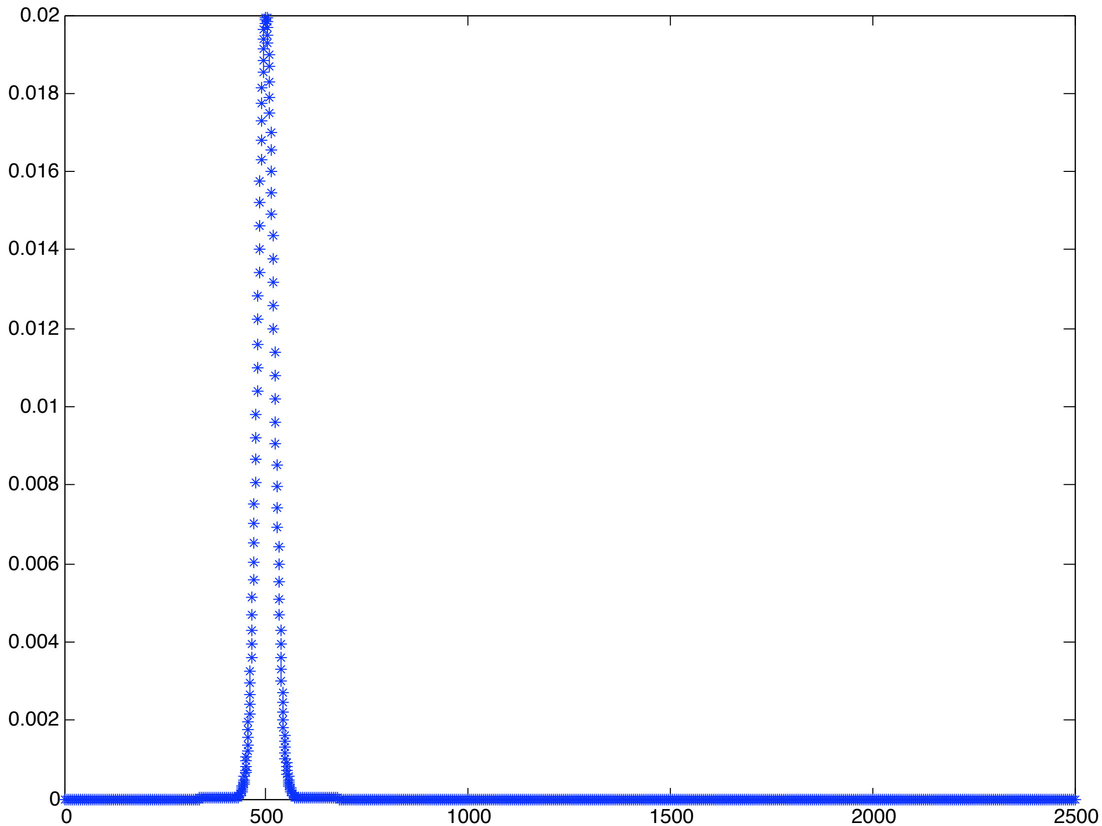
• **Comment:** Here we also know the standard deviation $\sigma = \sqrt{np(1-p)} = 20$.

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, $P(X \geq 540) \approx 0.0249 \;\; << \;\; 0.\overline{925}$.



Chebyshev inequality.

Theorem. (Chebyshev inequality.) If X is a random variable with finite mean μ 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

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

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 μ 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.

St. Petersburg paradox.

Suppose one plays a gambling game with probability of winning equal to the probability of losing. Think of tossing a fair coin. If the player bets M dollars on one of the two outcomes, then either the player wins another M dollars or looses the M dollars already at stake.

In the **St. Petersburg paradox** , the player begins with betting \$1. The strategy is to double the stake amount, betting \$2 in the second game, \$4 in the third game, \$8 in the fourth game, and so on. The player quits after the first win.