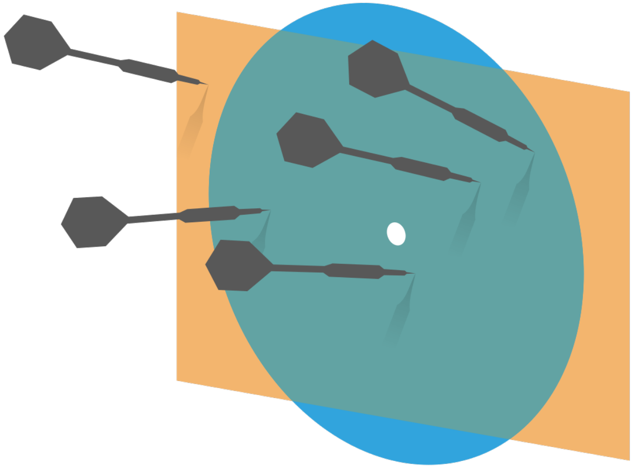
Probability and Statistics for Computer Science



Credit: wikipedia

Who discovered this?

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 02.25.2020

Last time

- Random Variable
  - *Review*
  - *The weak law of large numbers*

Proof of Weak law of large numbers

- Apply Chebyshev's inequality

$$P(|\overline{\mathbf{X}} - E[\overline{\mathbf{X}}]| \geq \epsilon) \leq \frac{var[\overline{\mathbf{X}}]}{\epsilon^2}$$

- Substitute $E[\overline{\mathbf{X}}] = E[X]$ and $var[\overline{\mathbf{X}}] = \frac{var[X]}{N}$

$$P(|\overline{\mathbf{X}} - E[\mathbf{X}]| \geq \epsilon) \leq \frac{var[\mathbf{X}]}{N\epsilon^2} \xrightarrow[N \to \infty]{} \mathbf{0}$$

$$\lim_{N\to\infty} P(|\overline{\mathbf{X}} - E[X]| \geq \epsilon) = 0$$

Applications of the Weak law of large numbers

- The law of large numbers ***justifies using simulations*** (instead of calculation) to estimate the expected values of random variables

$$\lim_{N \to \infty} P(|\overline{\mathbf{X}} - E[X]| \geq \epsilon) = 0$$

- The law of large numbers also ***justifies using histogram*** of large random samples to approximate the probability distribution function $P(x)$, see proof on

Pg. 353 of the textbook by DeGroot, et al.

Histogram of large random IID samples approximates the probability distribution

- The law of large numbers justifies using histograms to approximate the probability distribution. Given ***N*** IID random variables X_1, …, X_N
  - According to the law of large numbers

$$\overline{\mathbf{Y}} = \frac{\sum_{i=1}^{N} Y_i}{N} \xrightarrow{N \to \infty} E[Y_i]$$

  - As we know for indicator function

$$E[Y_i] = P(c_1 \leq X_i < c_2) = P(c_1 \leq X < c_2)$$

Probability using the property of Independence: Airline overbooking

- An airline has a flight with **s** seats. They always sell **t** (**t**>**s**) tickets for this flight. If ticket holders show up independently with probability **p**, what is the probability that the flight is overbooked ?

$$\text{P(overbooked)} = \sum_{u=s+1}^{t} C(t,u)p^u(1-p)^{t-u}$$

Simulation of airline overbooking

- An airline has a flight with **7** seats. They always sell 12 tickets for this flight. If ticket holders show up independently with probability **p**, estimate the following values
  - Expected value of the number of ticket holders who show up
  - Probability that the flight being overbooked
  - Expected value of the number of ticket holders who can’t fly due to the flight is overbooked.

Conditional expectation

- Expected value of X conditioned on event A:

$$E[X|A] = \sum_{x \in D(X)} xP(X = x|A)$$

- Expected value of the number of ticketholders not flying

$$E[NF|overbooked] = \sum_{u=s+1}^{t} (u-s) \frac{\binom{t}{u} p^u (1-p)^{t-u}}{\sum_{v=s+1}^{t} \binom{t}{v} p^v (1-p)^{t-v}}$$

Simulate the arrival

- Expected value of the number of ticket holders who show up

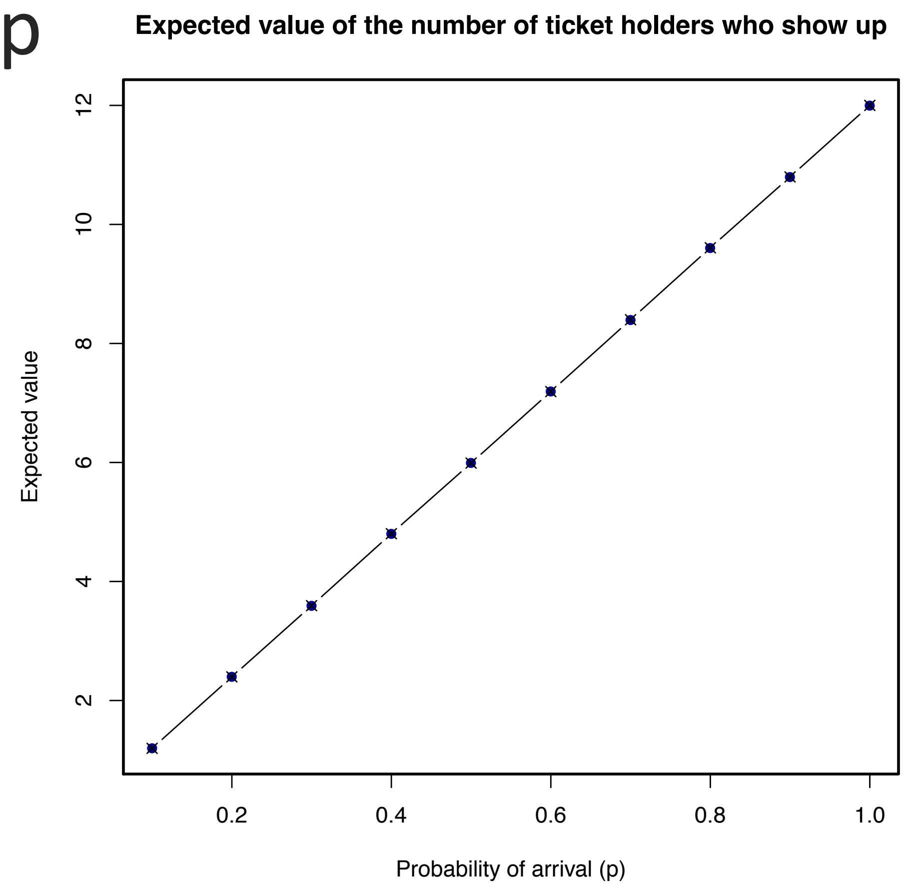
nt=100000, t= 12, s=7, p=0.1, 0.2, … 1.0

We generate a matrix of random numbers from uniform distribution in [0,1],
Any number < p is considered an arrival

Simulate the arrival

- Expected value of the number of ticket holders who show up

nt=100000, t= 12, s=7, p=0.1, 0.2, … 1.0



Simulate the expected probability of overbooking

- Expected probability of the flight being overbooked

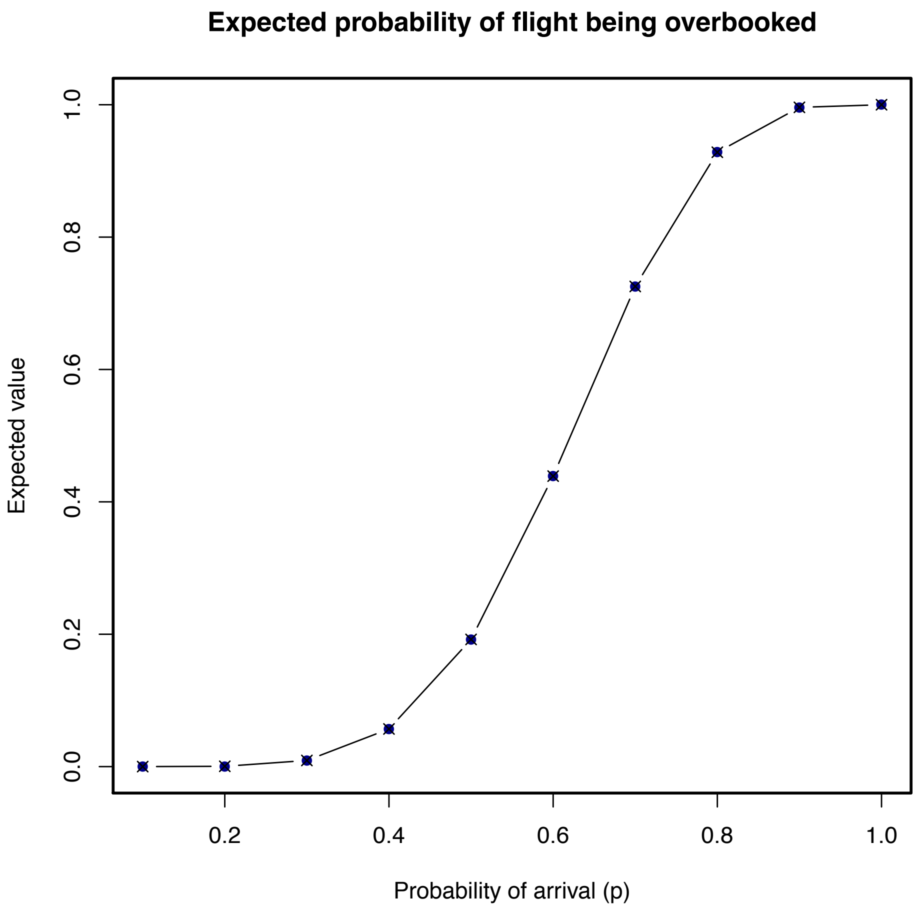
t= 12, s=7, p=0.1, 0.2, … 1.0

- **Expected probability** is equal to the **expected value of indicator function**. Whenever we have Num of arrival > Num of seats, we mark it with an indicator function. Then estimate with the sample mean of indicator functions.

Simulate the expected probability of overbooking

- Expected probability of the flight being overbooked

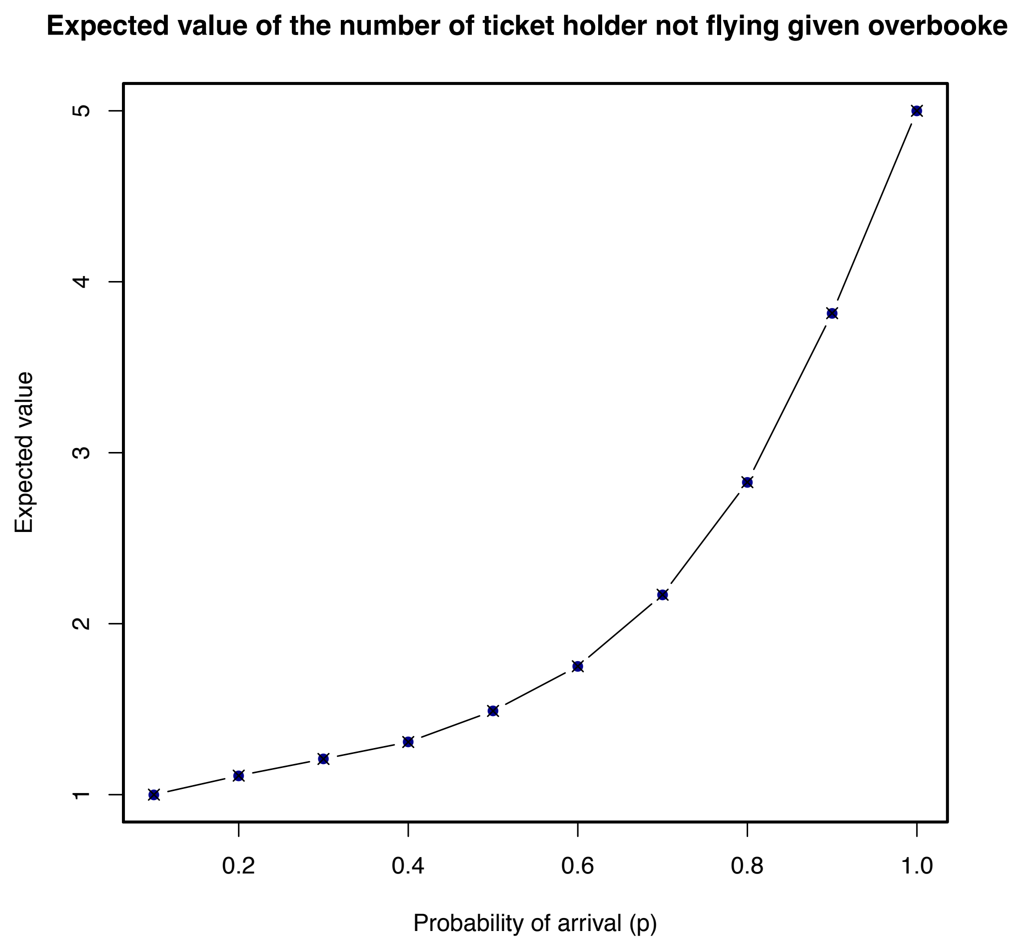
nt=100000,
t= 12, s=7,
p=0.1, 0.2, … 1.0



Simulate the expected value of the number of grounded ticket holders given overbooked

- Expected value of the number of ticket holders who can't fly due to the flight being overbooked

Nt=200000,
t= 12, s=7,
p=0.1, 0.2, … 1.0



Objectives

- Important known discrete probability distributions
- Continuous Random Variable

The classic discrete distributions

- Bernoulli
- Binomial
- Geometric
- Discrete uniform

Bernoulli distribution

- A random variable X is **Bernoulli** if it takes on two values 0 and 1 such that P(X=1) = p, P(X=0)=1-p

$$E[X] = p$$

$$var[X] = p(1-p)$$

Jacob Bernoulli (1654-1705)

Credit: wikipedia

Bernoulli distribution

- Examples
  - Tossing a biased (or fair) coin
  - Making a free throw
  - Rolling a six-sided die and checking if it shows 6
  - **Any indicator function** of a random variable

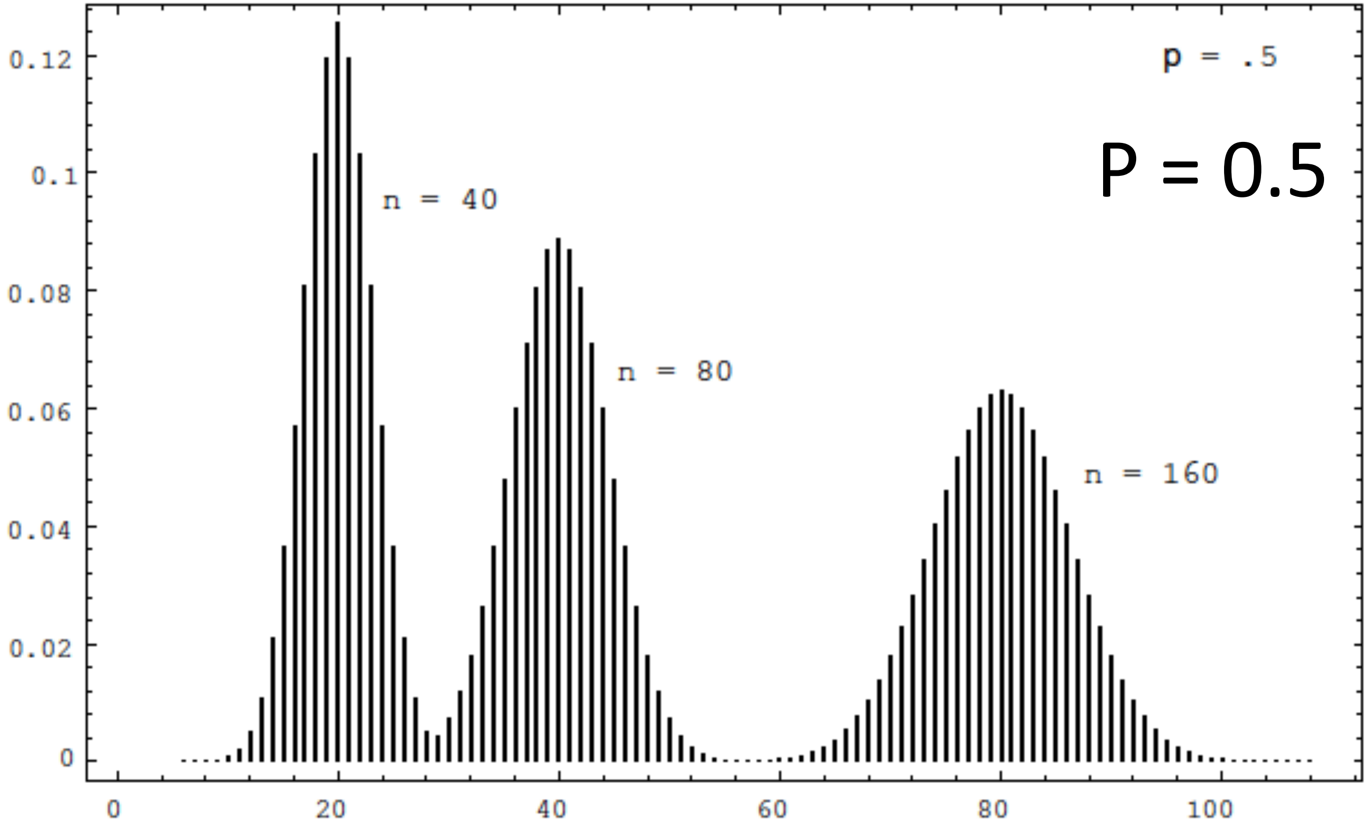
Binomial distribution

- The Galton Board

  http://www.randomservices.org/random/apps/GaltonBoardExperiment.html

- Remember the airline problem?

Binomial distribution



P = 0.5

Credit: Prof. Grinstead

Binomial distribution

- A discrete random variable X is binomial if

$$P(X = k) = \binom{N}{k} p^k (1-p)^{N-k} \quad for\ integer\ 0 \leq k \leq N$$

with $$E[X] = Np \quad \& \quad var[X] = Np(1-p)$$

- Examples
  - If we roll a six-sided die N times, how many sixes we will see
  - If I attempt **N** free throws, how many points will I score
  - **What is the sum of N independent and identically distributed Bernoulli trials?**

Expectations of Binomial distribution

- A discrete random variable X is binomial if

$$P(X = k) = \binom{N}{k} p^k (1-p)^{N-k} \quad for\ integer\ 0 \leq k \leq N$$

with $E[X] = Np \quad \& \quad var[X] = Np(1-p)$

Binomial distribution: die example

- Let X be the number of sixes in 36 rolls of a fair six-sided die. What is P(X=k) for k =5, 6, 7

- Calculate E[X] and var[X]

Geometric distribution

- A discrete random variable X is geometric if

$$P(X = k) = (1-p)^{k-1}p \qquad k \geq 1$$

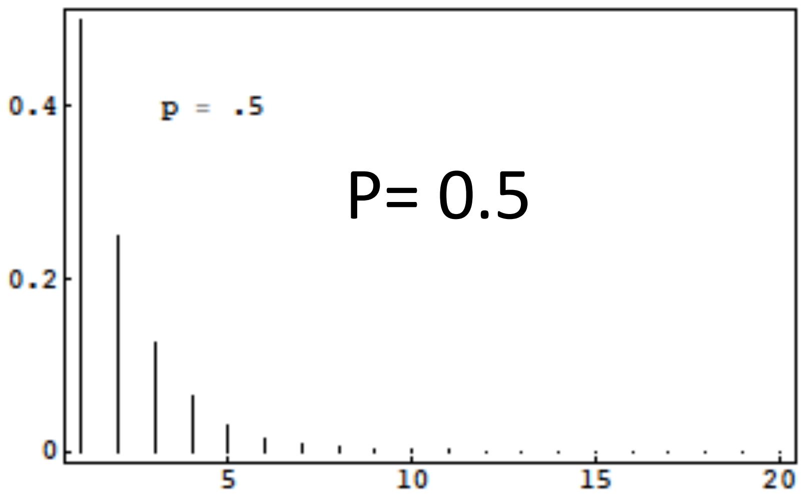
H, TH, TTH, TTTH, TTTTH, TTTTTH,...

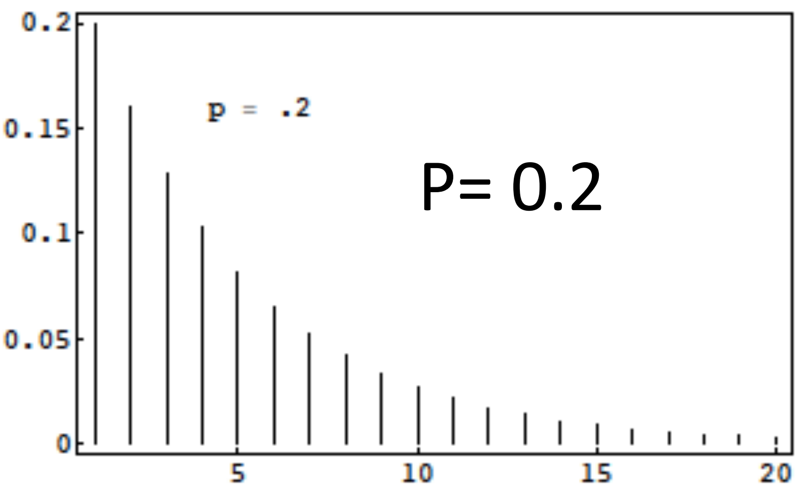
- Expected value and variance

$$E[X] = \frac{1}{p} \quad \& \quad var[X] = \frac{1-p}{p^2}$$

Geometric distribution

$$P(X = k) = (1-p)^{k-1}p \qquad k \geq 1$$





Credit: Prof. Grinstead

Geometric distribution

- Examples:
  - How many rolls of a six-sided die will it take to see the first 6?
  - How many Bernoulli trials must be done before the first 1?
  - How many experiments needed to have the first success?
  - Plays an important role in the **theory of queues**

Derivation of geometric expected value

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

Derivation of geometric expected value

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

$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

Derivation of geometric expected value

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

$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

$$= \frac{p}{1-p}\sum_{k=1}^{\infty} k(1-p)^{k}$$

Derivation of geometric expected value

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

$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

$$= \frac{p}{1-p}\sum_{k=1}^{\infty} k(1-p)^{k}$$

- For we have

this power series:

Derivation of geometric expected value

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

$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

$$= \frac{p}{1-p}\sum_{k=1}^{\infty} k(1-p)^{k}$$

- For we have this power series:

$$\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}; \quad |x| < 1$$

Derivation of geometric expected value

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

$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

$$= \frac{p}{1-p}\sum_{k=1}^{\infty} k(1-p)^{k}$$

$$x = 1 - p$$

- For we have

this power series:

$$\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}; \quad |x| < 1$$

Derivation of geometric expected value

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

$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

$$= \frac{p}{1-p}\sum_{k=1}^{\infty} k(1-p)^{k} = \frac{1}{p}$$

- For we have this power series:

$$\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}; \quad |x| < 1$$

Derivation of the power series

$$S(x) = \sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}; \quad |x| < 1$$

Proof:

$$\frac{S(x)}{x} = \sum_{n=1}^{\infty} nx^{n-1}\ ; \ \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}; \quad |x| < 1$$

$$\int_0^x \frac{S(t)}{t} = \sum_{n=1}^{\infty} x^n = x \cdot \frac{1}{1-x} = \frac{x}{1-x}$$

$$\frac{S(x)}{x} = \left(\frac{x}{1-x}\right)'$$

$$S(x) = \frac{x}{(1-x)^2}$$

Geometric distribution: die example

- Let X be the number of rolls of a fair six-sided die needed to see the first 6. What is $P(X = k)$ for k = 1, 2?

- Calculate E[X] and var[X]

$$E[X] = \frac{1}{p} \quad \& \quad var[X] = \frac{1-p}{p^2}$$

Betting brainteaser

- What would you rather bet on?
  - How many rolls of a fair six-sided die will it take to see the first 6?
  - How many sixes will appear in 36 rolls of a fair six-sided die?
- Why?

Multinomial distribution

- A discrete random variable X is Multinomial if

$$P(X_1 = n_1, X_2 = n_2, ..., X_k = n_k) = \frac{N!}{n_1!n_2!...n_k!} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}$$

$$where \;\; N = n_1 + n_2 + ... + n_k$$

- The event of throwing *N* times the k-sided die to see the probability of getting n_1 X_1, n_2 X_2, n_3 X_3...n_k X_k

Multinomial distribution

- A discrete random variable X is Multinomial if

$$P(X_1 = n_1, X_2 = n_2, ..., X_k = n_k) = \frac{N!}{n_1!n_2!...n_k!} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}$$

$$where \;\; N = n_1 + n_2 + ... + n_k$$

- The event of throwing k-sided die to see the probability of getting n_1 X_1, n_2 X_2, n_3 X_3...

ILLINOIS?

$$\frac{8!}{3!2!1!1!1!}$$

I L

Multinomial distribution

- Examples
  - If we roll a six-sided die N times, how many of each value will we see?
  - What are the counts of N independent and identical distributed trials?
  - This is very widely used in genetics

Multinomial distribution: die example

- What is the probability of seeing 1 one, 2 twos, 3 threes, 4 fours, 5 fives and 0 sixes in 15 rolls of a fair six-sided die?

Discrete uniform distribution

- A discrete random variable X is uniform if it takes k different values and

$$P(X = x_i) = \frac{1}{k}$$ For all x_i that X can take

- For example:
  - Rolling a fair k-sided die
  - Tossing a fair coin (k=2)

Discrete uniform distribution

- Expectation of a discrete random variable X that takes k different values uniformly

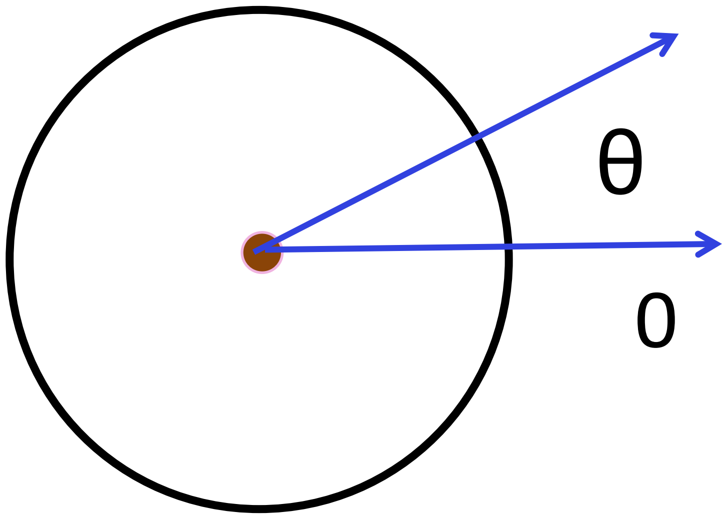
$$E[X] = \frac{1}{k}\sum_{i=1}^{k} x_i$$

- Variance of a uniformly distributed random variable X.

$$var[X] = \frac{1}{k}\sum_{i=1}^{k} (x_i - E[X])^2$$

Example of a continuous random variable

- The spinner



$$\theta \in (0, 2\pi]$$

- The sample space for all outcomes is not countable

Probability density function (pdf)

- For a continuous random variable X, the probability that X=x is essentially zero for all (or most) x, so we can't define $P(X = x)$
- Instead, we define the **probability density function** (pdf) over an infinitesimally small interval dx, $p(x)dx = P(X \in [x, x + dx])$
- For $a < b$ $$\int_a^b p(x)dx = P(X \in [a, b])$$

Properties of the probability density function

- $p(x)$ **resembles** the probability function of discrete random variables in that
  - $p(x) \geq 0$ for all x
  - The probability of X taking all possible values is 1.

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

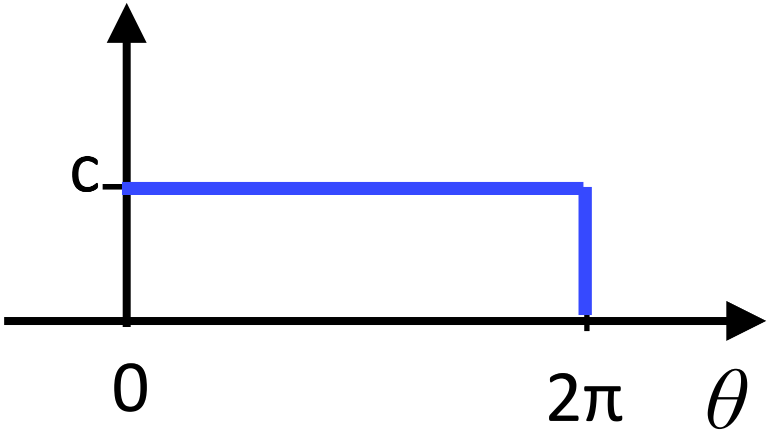
Properties of the probability density function

- $p(x)$ **differs** from the probability distribution function for a discrete random variable in that
  - $p(x)$ is not the probability that $X = x$
  - $p(x)$ can exceed 1

Probability density function: spinner

- Suppose the spinner has equal chance stopping at any position. What's the pdf of the angle θ of the spin position?

$$p(\theta) = \begin{cases} c & if\ \theta \in (0, 2\pi] \\ 0 & otherwise \end{cases}$$



- For this function to be a pdf,

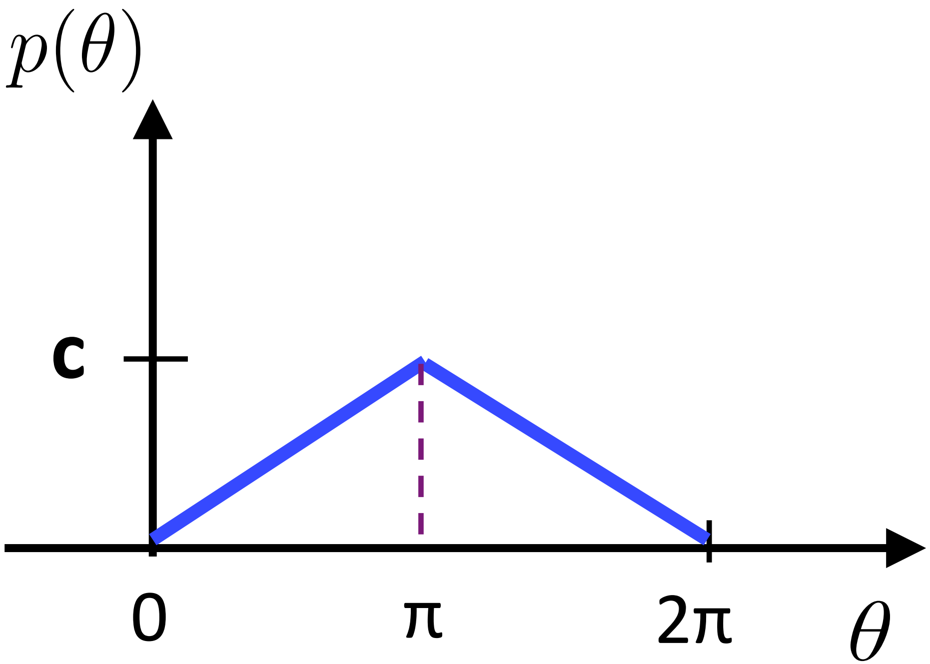
Then $$\int_{-\infty}^{\infty} p(\theta) d\theta = 1$$

Probability density function: spinner

- What the probability that the spin angle θ is within [$\frac{\pi}{12}, \frac{\pi}{7}$]?

Q: Probability density function: spinner

- What is the constant **c** given the spin angle θ has the following pdf?



A. 1
B. 1/π
C. 2/π
D. 4/π
E. 1/2π

Expectation of continuous variables

- Expected value of a continuous random variable X

$$E[X] = \int_{-\infty}^{\infty} x p(x) dx$$

weight

- Expected value of function of continuous random variable $Y = f(X)$

$$E[Y] = E[f(X)] = \int_{-\infty}^{\infty} f(x) p(x) dx$$

Probability density function: spinner

- Given the probability density of the spin angle θ

$$p(\theta) = \begin{cases} \frac{1}{2\pi} & if\ \theta \in (0, 2\pi] \\ 0 & otherwise \end{cases}$$

- The expected value of spin angle is

$$E[\theta] = \int_{-\infty}^{\infty} \theta p(\theta) d\theta$$

Properties of expectation of continuous random variables

- The linearity of expected value is true for continuous random variables.

$$\sum \longrightarrow \int$$

- And the other properties that we derived for variance and covariance also hold for continuous random variable

Q.

- Suppose a continuous variable has pdf

$$p(x) = \begin{cases} 2(1-x) & x \in [0,1] \\ 0 & otherwise \end{cases}$$

What is E[X]?

A. 1/2 B. 1/3 C. 1/4

D. 1 E. 2/3

$$E[X] = \int_{-\infty}^{\infty} x p(x) dx$$

Variance of a continuous variable

Assignments

- Work on Week5 material
- Next time: more classic known probability distributions

Additional References

- Charles M. Grinstead and J. Laurie Snell "Introduction to Probability”
- Morris H. Degroot and Mark J. Schervish "Probability and Statistics”

See you next time

See You!