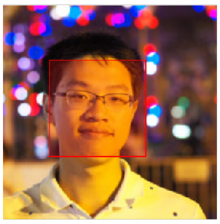
VE401 Recitation 1

About me

In[∘]:= `me =`  `;`

```
box = FindFaces[me];
HighlightImage[me, box]
```

Out[∘]=



- **Name**: GU Zhenhao
- **Cell Phone**: 13818175241
- **WeChat**: GaryGZH (or you can find me using the phone number above)
- **E-mail**: guzhenhao1@sjtu.edu.cn

- **Office Hour**: TBD
- **Recitation Class**: TBD

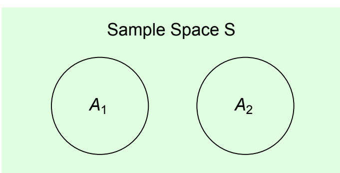
Contacting me via e-mail is preferred!

Introduction to Probability and Counting

Events and Mutually Exclusive Events

Intuition: ***Events*** are any subset of the sample space S. Events A_1 and A_2 are ***mutually exclusive*** if they cannot be true at the same time.

Out[229]=



Mathematical representation: $A_1 \cap A_2 = \emptyset$.

Cardano's Principle and Tree Diagrams

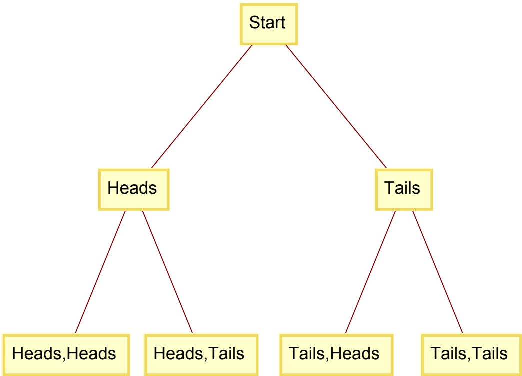
Intuition: The probability of an outcome A is the proportion of the ways that can lead to A, given that all the ways are equally likely and mutually exclusive.

Mathematical representation:

$$P[A] = \frac{\text{Number of ways leading to outcome } A}{\text{Number of ways experiment can proceed}}$$

Example: Tossing two (unbiased) coins, $P[\text{getting one head}] = \frac{2}{4} = 0.5$.

Out[∘]=



Counting

The number of ways leading to outcomes is calculated using the following: with a set $A = \{a_1, a_2, \ldots, a_n\}$ with n objects,

- Ways of picking an object k times, allowing repetition: n^k.
- Ways of choosing ordered k objects from A: $\frac{n!}{(n-k)!}$.
- Ways of choosing unordered k objects from A: $\frac{n!}{k!\,(n-k)!}$.
- Ways of partitioning A into k subsets, $A_1, \ldots, A_k$, with the number in the i^{th} subset is n_i: $\frac{n!}{n_1!\, n_2! \ldots n_k!}$.

Example:

Tossing a coin 4 times, what is the number of all possible outcomes?

$A = \{\text{heads}, \text{tails}\}$, $n = 2$. Total possible outcomes: $2^4 = 16$.

Rolling 10 four-sided dice, what is the number of ways of getting 3 ones, 2 twos, 4 threes and 1 four?

$A = \{1^{\text{st}}\text{ dice}, 2^{\text{nd}}\text{ dice}, \ldots, 10^{\text{th}}\text{ dice}\}$. We want to divide A into 4 subsets $A_1 :=$ dice with rolling result 1, $A_2 :=$ dice with rolling result 2, and same for A_3, A_4. In our case we need $|A_1| = 3$,

$|A_2| = 2$, $|A_3| = 4$, $|A_4| = 1$, then the number of total possible ways:

In[∘]:= $\dfrac{10!}{3!\,2!\,4!\,1!}$

Out[∘]= 12 600

Probability Measures & Spaces

Paraphrase: If a function P wants to be a ***probability measure***, it needs to

- map an event to its probability, which is in [0, 1],
- map the whole sample (event) space to 1, i.e. $P[S] = 1$,
- For mutually exclusive events $A_1, \ldots, A_k$, $P[A_1 \cup A_2 \cup \ldots \cup A_k] = P[A_1] + P[A_2] + \ldots + P[A_k]$.

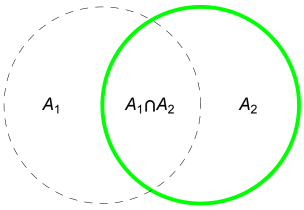
Properties:

- $P[S] = 1$,
- $P[\emptyset] = 0$,
- $P[S \setminus A] = 1 - P[A]$,
- $P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2]$.

Conditional Probability

Intuition: Given A_2 is true, the probability of A_1 is also true is the ***conditional probability*** $P[A_1 \mid A_2]$. We are basically finding where A_1 is true inside the A_2 circle.

Out[∘]=



Mathematical representation:

$$P[A_1 \mid A_2] = \frac{P[A_1 \cap A_2]}{P[A_2]}$$

$$P[A_1 \cap A_2] = P[A_1 \mid A_2]\, P[A_2] = P[A_2 \mid A_1]\, P[A_1]$$

Independence

Intuition: Events A and B are ***independent*** if outcome of A does not effect outcome of B, and vice versa.

Mathematical representation:

$$P[A \mid B] = P[A] \qquad \text{if } P[B] \neq 0$$

$$P[B \mid A] = P[B] \qquad \text{if } P[A] \neq 0$$

$$P[A \cap B] = P[A]\, P[B] \quad \text{(Why?)}$$

Summary: given events A and B, what happens to $P[A \cap B]$, $P[A \cup B]$, $P[A \mid B]$, $P[A \mid \neg B]$, and $P[\neg A \mid B]$? You are encouraged to fill out this table by yourself first.

A and B are ...	mutually exclusive	independent
$P[A \cap B]$		
$P[A \cup B]$		
$P[A \mid B]$		
$P[A \mid \neg B]$		
$P[\neg A \mid B]$		

What does this table imply?

A and B are ...	mutually exclusive	independent
$P[A \cap B]$	0	$P[A]\, P[B]$
$P[A \cup B]$	$P[A] + P[B]$	$P[A] + P[B] - P[A]\, P[B]$
$P[A \mid B]$	0	$P[A]$
$P[A \mid \neg B]$	$\frac{P[A]}{1-P[B]}$	$P[A]$
$P[\neg A \mid B]$	1	$1 - P[A]$

This table means if events of zero probability are excluded, mutually exclusive events are not independent and independent events are not mutually exclusive.

Total Probability

Purpose: To write down a probability using the sum of conditional probabilities.

Mathematical representation:

$$P[B] = \sum_{j=1}^{n} P\left[B \mid A_j\right] P\left[A_j\right] \qquad \text{if } A_1 \ldots A_n \subset S \text{ are mutually exclusive and } A_1 \cup \ldots \cup A_n = S$$

Bayes's Theorem

Purpose: To switch sides of a conditional probability, i.e. from $P[A \mid B]$ to $P[B \mid A]$.

Mathematical representation:

$$P[A_k \mid B] = \frac{P[B \cap A_k]}{P[B]} = \frac{P[B \mid A_k]\, P[A_k]}{P[B \mid A_k]\, P[A_k] + P[B \mid \neg A_k]\, P[\neg A_k]}$$

Furthermore, if $A_1 \ldots A_n \subset S$ are mutually exclusive and $A_1 \cup \ldots \cup A_n = S$,

$$P[A_k \mid B] = \frac{P[B \mid A_k]\, P[A_k]}{\sum_{j=1}^{n} P[B \mid A_j]\, P[A_j]}$$

Example: I will recommend this video https://www.youtube.com/watch?v=HZGCoVF3YvM, which gives a nice visualization of Bayes's theorem.

Discrete Random Variable

Discrete Random Variable and Probability Density Function (PDF)

Paraphrase: a ***discrete random variable*** X maps the sample space to a countable subset Ω of $\mathbb{R}$, with each number representing an event, and the ***probability density function*** f_X maps the subset to its probability. The PDF must follow

- $f_X(x) > 0$ for all x.
- $\sum_{x\in\Omega} f_X(x) = 1$.

Note: For various distribution of random variable X, we need to know its

- parameter(s),
- $\mathrm{E}[X]$,
- $\mathrm{Var}[X]$,
- probability density function (PDF),
- cumulative distribution function (CDF),
- moment generating function (MGF), and
- when to use it?

Expectation

Intuition: given a set of data, what would I expect the next number to be?

Mathematical representation: for discrete random variable, $\mathrm{E}[X] = \mu_X = \mu = \sum_{x\in\Omega} x\, f_X(x)$.

Properties: for any random variable X and Y,

- For a constant $c \in \mathbb{R}$, $\mathrm{E}[c] = c$, $\mathrm{E}[c\,X] = c\,\mathrm{E}[X]$,
- $\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$,
- For any function $\varphi : \Omega \to \mathbb{R}$, $\mathrm{E}[\varphi \circ X] = \sum_{x\in\Omega} \varphi(x)\, f_X(x)$.

What do these properties imply?

$\mathrm{E}[\cdot]$ is a linear operation!

Variance and Standard Variance

Intuition: how much does random variable deviate from the mean?

Mathematical representation:

- Variance: $\text{Var}[X] = {\sigma_X}^2 = \sigma^2 = \text{E}\left[\left(X - \text{E}[X]\right)\right]^2 = \text{E}\left[X^2\right] - \text{E}[X]^2$.
- Standard variance: $\sigma_X = \sqrt{\text{Var}[X]}$.

Properties:

- For a constant $c \in \mathbb{R}$, $\text{Var}[c] = 0$, $\text{Var}[c\,X] = c^2\,\text{Var}[X]$.
- For independent random variable X and Y, $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$.

Moment and Moment Generating Function (MGF)

Intuition: MGF encodes the sequence of all ***moments***, $\text{E}\left[X^0\right]$, $\text{E}\left[X^1\right]$, $\text{E}\left[X^2\right]$, ... into one function.

Mathematical representation: MGF exists iff $\text{E}\left[e^{t\,X}\right]$ exists, in which case

$$m_X(t) = \sum_{k=0}^{\infty} \frac{\text{E}\left[X^k\right]}{k!}\,t^k = \text{E}\left[e^{t\,X}\right]$$

and the k^{th} moment can be calculated using $\text{E}\left[X^k\right] = \left(\frac{d^k\,m_X(t)}{d\,t^k}\right)_{t=0}$.

Properties: Assume MGF exists,

- if two distributions have same MGF, then two distributions are identical.
- For any constant $\alpha, \beta \in \mathbb{R}$, $m_{\alpha\,X+\beta}(t) = e^{\beta\,t}\,m_X(\alpha\,t)$.
- For independent random variable X_1, X_2, $m_{X_1+X_2} = m_{X_1}\,m_{X_2}$.

Cumulative Distribution Function (CDF)

Mathematical representation: $F_X(x) := P[X \leq x] = \sum_{y \leq x} f_X(y)$.

Bernoulli Distribution

Purpose: the probability of success in 1 trial?

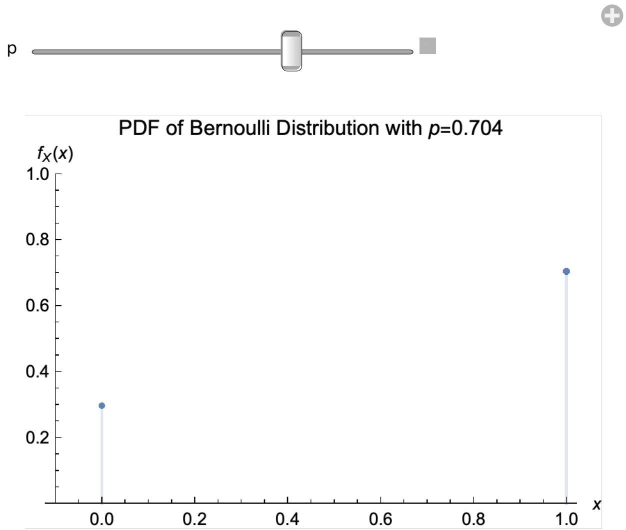
Parameter and properties:

- $p \in [0, 1]$, describing the probability of success. $q := 1 - p$.

Mean	Variance	PDF	CDF	MGF

p	$p\,q$	$\begin{cases} q & x == 0 \\ p & x == 1 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 0 & x < 0 \\ q & 0 \le x < 1 \\ 1 & \text{otherwise} \end{cases}$	$q + e^t\,p$

Out[141]=



Binomial Distribution

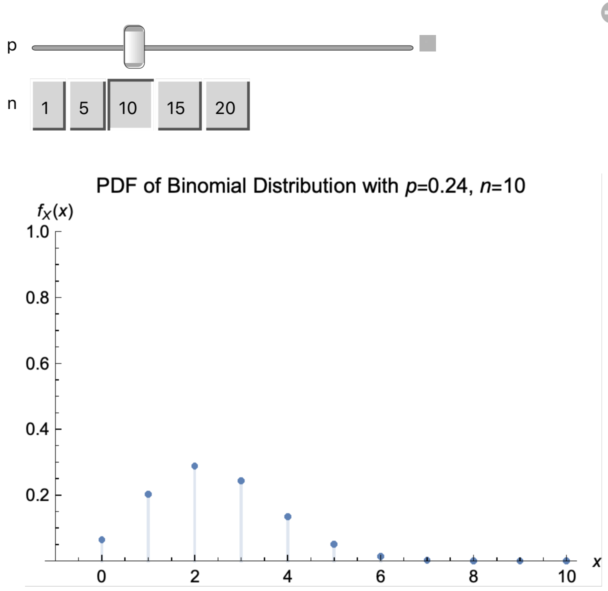
Purpose: how many successes in n trial(s)?

Parameter and properties:

- $p \in [0, 1]$, describing the probability of success in each trial. $q := 1 - p$.
- $n \in \{0, 1, 2, \ldots\}$ is the number of trials.

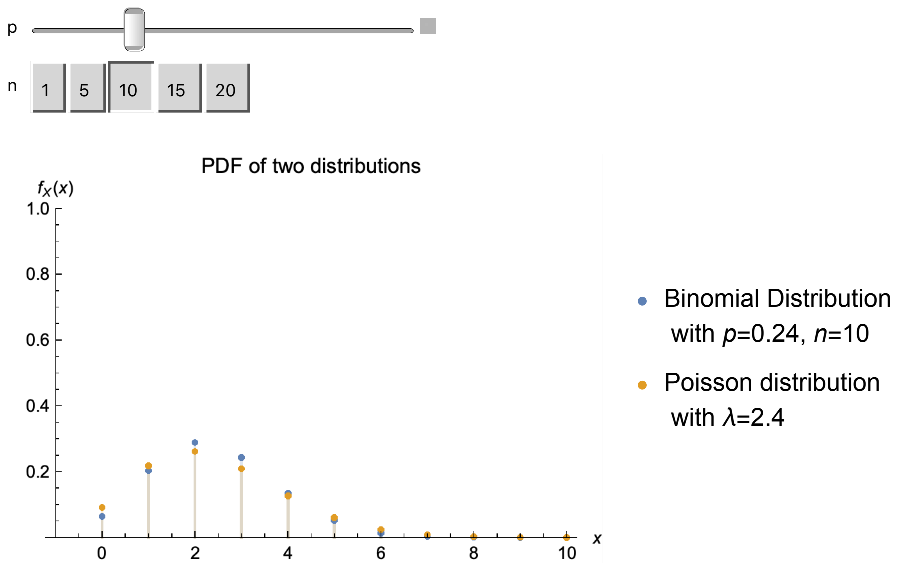
Mean	Variance	PDF	MGF
$n\,p$	$n\,p\,q$	$\begin{cases} \binom{n}{x} p^x\, q^{n-x} & 0 \le x \le n \\ 0 & \text{otherwise} \end{cases}$	$\left(p\left(e^t - 1\right) + 1\right)^n$

Out[140]=



Approximation of PDF: if n is large and p is small, the PDF can be approximated by Poisson distribution with parameter $\lambda = n\,p$.

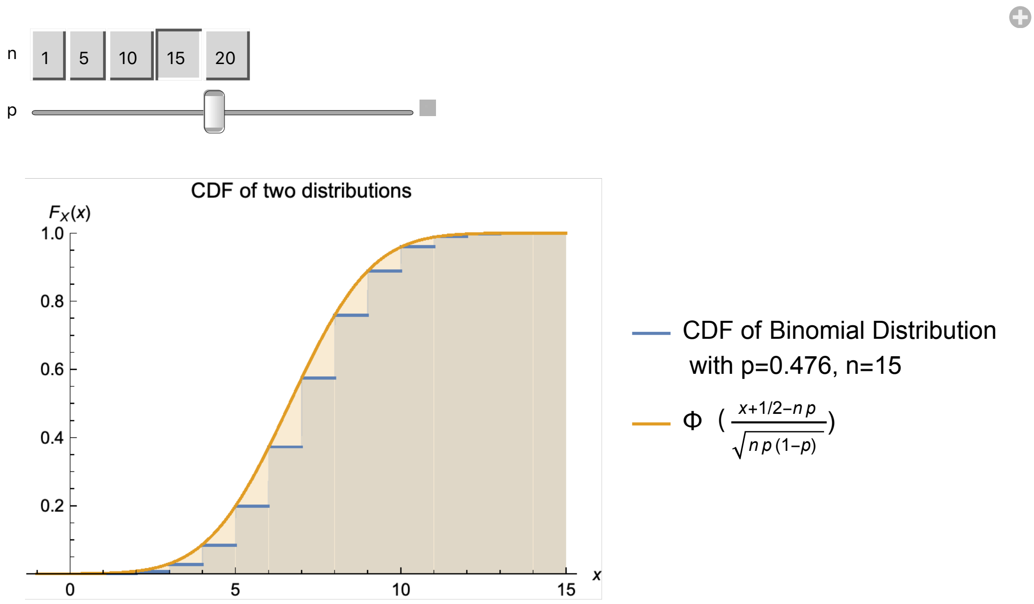
Out[137]=



Approximation of CDF: if $\begin{cases} n > 10 & \text{if } p \text{ is close to } 1/2 \\ n\,p > 5 & \text{if } p \le 1/2 \\ n(1-p) > 5 & \text{if } p > 1/2 \end{cases}$, the CDF at $y \in \mathbb{N}$ can be approximated by normal distribution.

$$P[X \le y] = \sum_{x=0}^{y} \binom{n}{x} p^x (1-p)^{n-x} \approx \Phi\left(\frac{y+1/2-n\,p}{\sqrt{n\,p(1-p)}}\right)$$

Out[231]=



Geometric Distribution

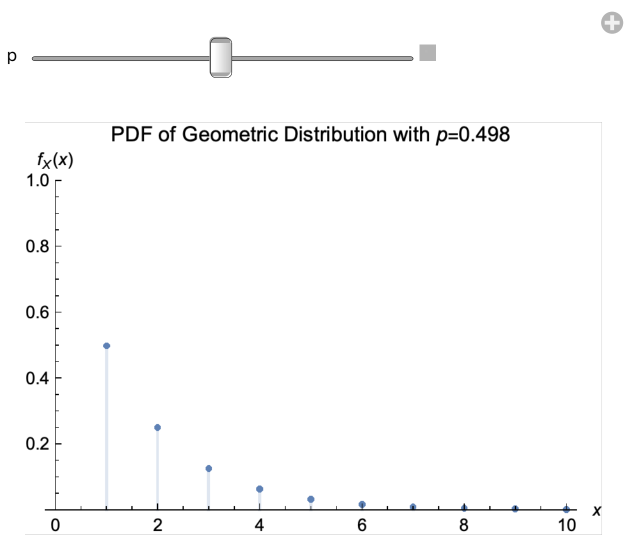
Purpose: how many trials until first success?

Parameter and properties:

- $p \in [0, 1]$, describing the probability of success in each trial. $q := 1 - p$.
- $n \in \{0, 1, 2, \ldots\}$ is the number of trials.

Mean	Variance	PDF	CDF	MGF
$\frac{1}{p}$	$\frac{q}{p^2}$	$\begin{cases} q^{x-1}\,p & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 - q^{\lfloor x \rfloor} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$	$\frac{p\,e^t}{1-q\,e^t}$

Out[136]=



Hypergeometric Distribution

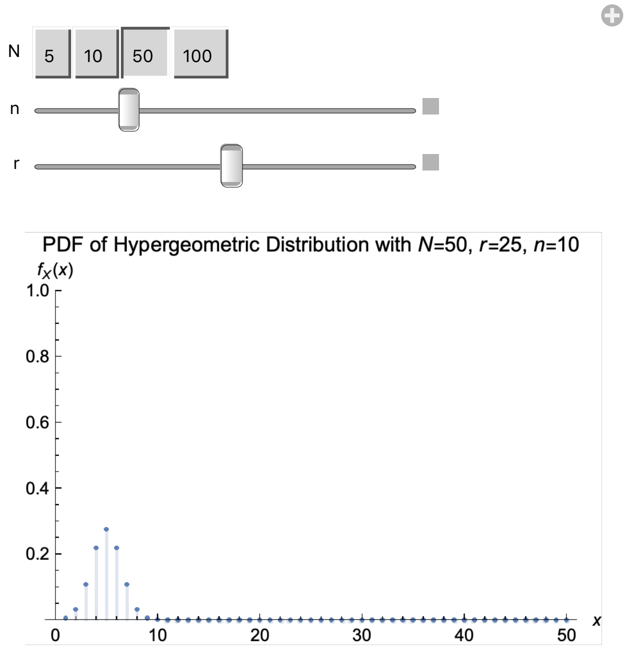
Purpose: Select n samples from N objects (within which r objects have trait), what is the probability of having x objects with trait?

Parameter and properties:

- $N \in \{0, 1, 2, \ldots\}$ is the number of total objects.
- $n \in \{0, 1, 2, \ldots, N\}$ is the number of samples.
- $r \in \{0, 1, 2, \ldots, N\}$ is the number of objects with traits. $p := \frac{r}{N}$, $q := 1 - p$.

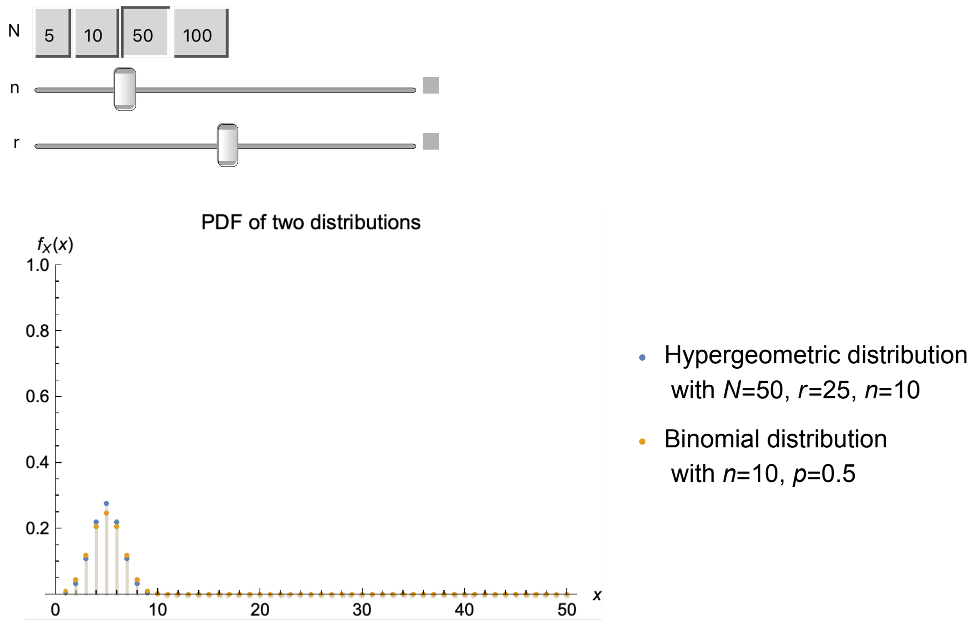
Mean	Variance	PDF
$n\,p$	$n\,p\,q\,\frac{N-n}{N-1}$	$\begin{cases} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} & 0 \le x \le N \\ 0 & \text{otherwise} \end{cases}$

Out[138]=



Approximation: if n / N is sufficiently small, it can be approximated by binomial distribution with parameter n and $p = r / N$.

Out[225]=



Poisson Distribution

Purpose: how many arrivals in a small time interval?

Parameter and properties:

- $\lambda > 0$, describing the rate of arrival.

Mean	Variance	PDF	MGF
λ	λ	$\begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$e^{\lambda(e^t-1)}$

Out[133]=