Mathematical Methods for Automation Engineering M
– *Discrete Probability Distributions* –

Andrea Mentrelli

Department of Mathematics &
Alma Mater Research Center on Applied Mathematics, AM^2
University of Bologna

andrea.mentrelli@unibo.it

A.Y. 2024/25

Discrete Probability Distributions

Probability distributions are classified in two families, according to the discrete/continuous nature of the random variables

We discuss some of the most important probability distributions of both families

- **Discrete probability distributions**: Bernoulli, binomial, geometric, Poisson, discrete uniform distributions
- **Continuous probability distributions**: continuous uniform, exponential, Rayleigh, gamma, Erlang, Weibull, Gaussian (normal) distributions

Discrete Probability Distributions

Definition (Bernoulli experiment)

We call *Bernoulli experiment* a random experiment with only two predictable outcomes, conventionally named **success** and **failure**

In the context of a Bernoulli experiment, the sample space Ω is partitioned in two events, i.e. $\Omega = \{E, \overline{E}\}$

- E: **success**, with probability p
- $\overline{E}$: **failure**, with probability $q = 1 - p$

The probability p is called **Bernoulli parameter** $(0 \leqslant p \leqslant 1)$

We shall introduce a random variable $X : \Omega \to \{0, 1\}$ that takes the values

- "1" in case of success (the event E occurs)
- "0" in case of failure (the event $\overline{E}$ occurs)

Definition (Bernoulli random variable)

Let $\boxed{X \text{ be a random variable}}$ which can only take the values "1" and "0"

$$\boxed{X : \Omega \to \{0, 1\}}$$

a set of value
0 AND 1

Such a random variable is called *Bernoulli random variable* and it is denoted as follows

$$X \sim Be(p)$$

"distributed as"

being p the probability associated to the value "1" (*success*)

$$p_X(k) = \mathcal{P}(X = k) = \begin{cases} p & (k = 1) \\ 1 - p & (k = 0) \end{cases} \Rightarrow \boxed{p_X(k) = p^k \, (1 - p)^{1-k}}$$

success

failure

Discrete Probability Distributions

Example

In the experiment "rolling a die" we define

- success: the event "the outcome is less than 3" $S \equiv \{\omega_1, \omega_2\}$ $2/6$
- failure: the event "the outcome is greater than or equal to 3" $F \equiv \{\omega_3, \omega_4, \omega_5, \omega_6\}$ $4/6$

This experiment is a Bernoulli experiment of parameter $p = 1/3$. The random variable $X \sim Be(p)$ and the corresponding probability mass function are

success

$p = 1/3$

$$X : \{F, S\} \to \{0, 1\} \qquad \begin{cases} X(F) = 0 \\ X(S) = 1 \end{cases}$$

$X \sim Be(1/3)$

$$p_X(k) = \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{1-k} = \frac{2^{1-k}}{3}$$

with $k \in \{0, 1\}$

Theorem (Initial moments of the Bernoulli random variable)

The initial moment of order s of the Bernoulli random variable is

$$\alpha_s(X) = p$$

$$\boxed{P_X(x_i) = p_i}$$

Proof. Applying the definition of initial moment of order s

$$\alpha_s(X) = \sum_{i=1}^{2} x_i^s \, p_i = 0^s \, (1-p) + 1^s \, p = p$$

$$\sum_{i=1}^{2} x_i^s$$

□

The initial moments α_s of a Bernoulli random variable do not depend on the order s

Theorem (Expected value and Variance of the Bernoulli r. v.)

Expected value and variance of the Bernoulli random variable are

$$\boxed{\mu_X = p} \qquad \boxed{\sigma_X^2 = p\,(1-p) = p\,q}$$

Proof. Applying the definition

$$\sigma_X^2 = \sum_{i=1}^{2} (x_i - \mu_X)^2\, p_i = (-p)^2\,(1-p) + (1-p)^2\, p = p\,(1-p) = pq$$

The variance can also be calculated by means of the well-known theorem

$$\sigma_X^2 = E\left(X^2\right) - E\left(X\right)^2 = p - p^2 = p\,(1-p) = p\,q$$

□

Discrete Probability Distributions

Example

Let us consider the experiment "tossing of two fair dice". Let X be the random variable that takes the value "1" when the two dice show the same points (success), "0" otherwise (failure)

1. The probability mass function $p_X(x)$ is

$$p_X(X \overset{\text{fail}}{=} 0) = \frac{5}{6} \qquad p_X(X \overset{\text{success}}{=} 1) = \frac{1}{6}$$

2. Expected value and variance of the random variable X are

$$\mu_X = p = \frac{1}{6} \qquad \sigma_X^2 = p\,q = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$$

Discrete Probability Distributions

Definition (Bernoulli process)

A sequence of independent Bernoulli experiments, all with the same parameter p, is called *Bernoulli process*

- **sequence of repetitions**: definite or indefinite repetitions of the experiment
- **independent repetitions**: the outcome (thus, the probability) of a repetition is not affected by the outcomes of other repetitions (independent repetitions are called **trials**)

Examples of Bernoulli processes

- indefinite flipping of a coin until the outcome is *heads* ("success")
- twenty rolls of a die, counting the numbers of trials whose outcome is "3" ("success")
- check of n devices, counting the failed ones ("success")

Discrete Probability Distributions

There are several probability distributions frequently encountered in the applications which arise in the context of Bernoulli processes

Some of the most important are

- *binomial* distribution
- *geometric* distribution

Definition (Binomial random variable)

Given a Bernoulli process of parameter p made of n trials of an experiment, the *binomial random variable* X *of parameters* n *and* p

$$\boxed{X \sim B(n, p)}$$

is defined as the random variable that counts the number of "successes" in the $\boxed{n \text{ trials}}$

$$\boxed{X : \Omega \to \{0, 1, 2, \dots n\}}$$

- A binomial random variable of parameters $(1, p)$ is just a Bernoulli random variable

$$B(1, p) \equiv Be(p)$$

n = 1

- A binomial random variable of parameters n and p is the **sum** of n independent Bernoulli random variables of parameter p [more on this later] (say X_1 and X_2 are independent from each other)

Discrete Probability Distributions

Theorem (Bernoulli formula)

what are the probability for each X_k?

The probability of obtaining in a Bernoulli process (of $\boxed{n \text{ trials}}$) $\boxed{k \text{ successes}}$ and $\boxed{n-k \text{ failures}}$ is

trials

success

failures

$$P_{k,n} = \mathcal{P}(X = k) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \quad (k = 0, 1, \dots n)$$

failure prob.

success prob.

where $\boxed{p}$ is the probability of $\boxed{\text{success}}$ and $\boxed{q = 1 - p}$

Proof. Given a sequence of n trials, the number of possible outcomes with k successes and $n-k$ failures is the number of combinations of order k of n objects, i.e. $\binom{n}{k}$. Each of these outcomes has probability $p^k q^{n-k}$ □

Theorem (Expected value and Variance of the Binomial r. v.)

Expected value and the variance of the binomial random variable are

$$\boxed{\mu_X = np} \qquad \boxed{\sigma_X^2 = np(1-p) \equiv npq}$$

it's not general

What is the probability of $P(X = k)$ in n trials?

Let's say the random experiment outcomes are :

$$\overbrace{1\ 1\ 1\ 1\ 1}^{k}\ \overbrace{0\ 0\ 0\ 0\ 0\ 0}^{n-k} \longrightarrow p^k (1-p)^{n-k}$$

$$1\ 1\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 0$$

$$0\ 1\ 0\ 0\ 1\ 1\ 1\ 1\ 0\ 0\ 0$$

$\vdots$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Since all the probabilities are similar, it represents the combination of $\binom{n}{k}$

$$= \frac{n!}{k!(n-k)!}$$

Example (Statistical multiplexing)

An ISP (Internet Service Provider) offers 100 Mbps bandwidth, which are allocated to the customers in chunks of 5 Mbps. Let us assume that the customers are actively using the bandwidth 50% of the time (on average).

If the whole bandwidth is assigned to 20 customers, what is the probability that at a given time the bandwidth is fully utilized?

$n = 20$, k?

$$\mathcal{P}(X = 20) = P_{20,20} = \binom{20}{20}\left(\frac{1}{2}\right)^{20}\left(\frac{1}{2}\right)^{20-20} = \frac{1}{2^{20}} \simeq 9.53 \cdot 10^{-7}$$

If the bandwidth is allocated to 30 customers, what is the probability of not having more than 20 of them trying to simultaneously utilize the bandwidth? $n = 30$

$$\mathcal{P}(X \leqslant 20) = \sum_{k=0}^{20} P_{k,30} = \sum_{k=0}^{20}\binom{30}{k}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{30-k} \simeq 0.979$$

Definition (Geometric random variable) we don't know how much the repetition will be

Given an indefinite Bernoulli process of parameter p, we call *geometric random variable X of parameter p*, denoted by the symbol

$$\boxed{X \sim G(p)}$$

the random variable that counts the number of necessary trials before obtaining the first success

$$\boxed{X : \Omega \to \mathbb{N}}$$

including the last trials

The geometric random variable finds application in statistical sampling, when trials are repeated until a specific result is obtained. For example: testing electical components until a failed component is found

Discrete Probability Distributions

Theorem (Probability mass function of the geometric r. v.)

Given a geometric random variable $X \sim G(p)$, its probability mass function is

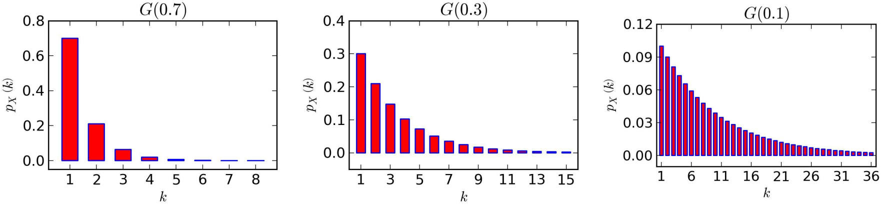
$$\underbrace{(1-p)(1-p)\dots(1-p)}_{k-1}\, p$$

$$p_X(k) \equiv \mathcal{P}(X = k) = \boxed{p\,(1-p)^{k-1}} \qquad (k = 1, 2, 3, \dots)$$

or, with a slightly different (but equivalent) formulation

$$p_X(k) \equiv \mathcal{P}(X = k) = p\,(1-p)^{k} \qquad (k = 0, 1, 2, \dots)$$

Discrete Probability Distributions



Remark. The geometric random variable is the only discrete random variable **without memory**: the probability of success does not depend on the "history" (What does that mean?) does not depend on previous results, or in the other word: no matter the results of previous experiments

Imposing

$$\mathcal{P}(X \geqslant s + t \mid X > t) = \mathcal{P}(X \geqslant s)$$

we find that

$$\mathcal{P}(X = 0) = p$$

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

Theorem (Expected value and Variance of the geometric r. v.)

The expected value and variance of the geometric random variable $X \sim G(p)$ *are*

$$\boxed{\mu_X = \frac{1}{p}} \qquad \boxed{\sigma_X^2 = \frac{q}{p^2}}$$

Proof. (Expected value) Recalling that $p_k = pq^{k-1}$ $(k = 1, 2, 3, \ldots)$, the expected value is computed as follows

$$\mu_X = \sum_{k=1}^{\infty} x_k p_k = \sum_{k=1}^{\infty} k\, pq^{k-1} = p \sum_{k=1}^{\infty} k\, q^{k-1} = p \frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) =$$

$$p \frac{d}{dq}\left(\frac{q}{1-q}\right) = p \frac{1}{(1-q)^2} = \frac{1}{p}$$

where it has been taken into account that, for $|\, q \,| < 1$, the following holds (geometric series): $\sum_{k=1}^{\infty} q^k = \frac{q}{1-q}$ □

Towards the Poisson distribution

Let us consider the probability of random events in **different contexts**

- The **molecules of a pollutant** are *randomly* dispersed. What is the probability of finding k such molecules in a fixed 1 cm^3?
  - n, number of molecules of pollutant → unknown
  - p, probability of finding a molecule in 1 cm^3 → unknown
  - a, average number of molecules of pollutant in 1 cm^3 → computable
- **Typos** are *randomly* distributed over the pages of a printed book. What is the probability of finding k typos in a given page?
  - n, number of typos → unknown
  - p, probability of finding a typo in a given page → unknown
  - a, average number of typos per page → computable
- **Radioactive decay** takes place *randomly* in time. What is the probability of detecting k decays on a given Δt?
  - n, number of atoms of a radioactive substance → unknown
  - p, probability of decay of a given atom in Δt → unknown
  - a, average number of decays in a given Δt → computable

Discrete Probability Distributions

Another example:

- **Car drivers** arrive *randomly* at the motorway barrier. What is the probability of counting k cars passing the barrier in a given day?
  - n, number of car drivers on the motorway
  - p, probability that a given driver passes the barrier that day
  - a, average number of drivers per day passing the barrier

All the previous examples have **common traits**

$n \gg$, $p \ll$, a is finite

1. a **large** number n of *"particles"* (molecules, typos, atoms, car drivers) are randomly dispersed in a *continuum* (temporal, spatial, or other)
2. the (independent) particles have a **small** probability p (the same for all the particles) of being found in the *unit* of the continuum
3. typically, the values of n and p are unknown, but the *finite* **average number** a of particles in the unit of the continuum is known
4. The number of particles in the unit of the continuum is the **random variable** of interest

- Let us define the event E_k as "the k^{th} driver passes the barrier" (probability p). Since p is the same for every car driver, **the events E_i are independent** and...
- ...The passing of each car driver at the barrier on a given day is a **Bernoulli experiment** of parameter p ("success" = the driver drives the motorway). Therefore, the process under study is thus a **Bernoulli process** $P(X=k)$
- The number X of cars driving the motorway corresponds to the successes in a Bernoulli process made of n trials of parameter $p \Rightarrow$ **binomial distribution** $X \sim B(n, p)$
- Typically, n and p are unknown, but a is given

$$\mu_X = \underset{\text{finite}}{\overset{\gg\;\;\ll}{np}} = a \rightarrow \text{known, finite} \qquad \Rightarrow \qquad p = a/n$$

then

unknown

$$p_X(k) = \mathcal{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \left(\frac{a}{n}\right)^k \left(1 - \frac{a}{n}\right)^{n-k}$$

$$
\begin{aligned}
p_X(k) &= \binom{n}{k} \left(\frac{a}{n}\right)^k \left(1-\frac{a}{n}\right)^{n-k} \\
&= \overbrace{\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}}^{\binom{n}{k}} \left(\frac{a}{n}\right)^k \left(1-\frac{a}{n}\right)^{n-k} \\
&= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right)}{k!}\, a^k \left(1-\frac{a}{n}\right)^{n-k}
\end{aligned}
$$

- $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right) \xrightarrow{n\to\infty} 1$
- $n \gg \quad \Rightarrow \quad \left(1-\frac{a}{n}\right)^{n-k} \simeq \left(1-\frac{a}{n}\right)^{n}$
- $e^x = 1 + x + \mathcal{O}\left(x^2\right) \;\Rightarrow\; e^{-x} \simeq 1 - x \;\Rightarrow\; \left(1-\frac{a}{n}\right) \simeq e^{-\frac{a}{n}} \;\Rightarrow\; \left(1-\frac{a}{n}\right)^n \simeq e^{-a}$

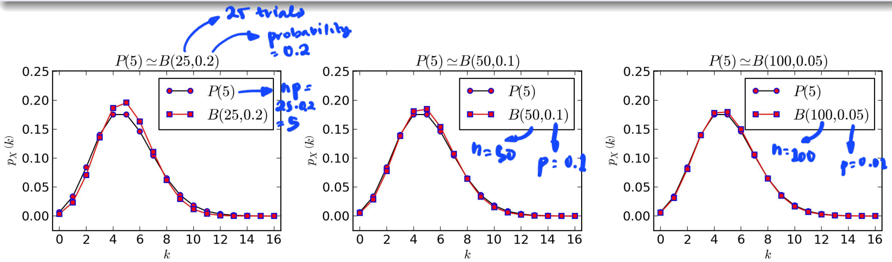
Finally (**Poisson approximation**):

$$
p_X(k) = \binom{n}{k} \left(\frac{a}{n}\right)^k \left(1-\frac{a}{n}\right)^{n-k} \xrightarrow{n\to\infty} \frac{a^k}{k!}\, e^{-a}
$$

Theorem (Poisson approximation) with limit

The binomial distribution of parameters (n, p) when $n \to \infty$, $p \to 0$ (with $np = a$ finite) tends to the Poisson formulae

$$\boxed{\lim_{\substack{n\to\infty \\ p\to 0 \\ np=a}} \binom{n}{k} p^k q^{n-k} = \frac{a^k}{k!} e^{-a}} \quad \Rightarrow \quad B(n,p) \sim P(a)$$



Discrete Probability Distributions

Definition (Poisson random variable)

Let X be a discrete random variable taking only non negative integer values ($X \in \mathbb{N}^+$). The random variable X is a *Poisson random variable of parameter $a > 0$*, denoted by the symbol

$$X \sim P(a)$$

if the probability that X takes on the value k is given by

$$p_X(k) = \mathcal{P}(X = k) = \frac{a^k}{k!} e^{-a} \qquad (k \geqslant 0)$$

$p_X(k)$

The Poisson random variable is well-defined. This is easily seen as follows

$=1$

- $\mathcal{P}(X = k) \geqslant 0$
- $\boxed{\sum_{k=0}^{\infty} \mathcal{P}(X = k)} = \sum_{k=0}^{\infty} \frac{a^k}{k!} e^{-a} = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} = e^{-a} e^{a} = 1$

Theorem (Expected value and Variance of the Poisson r. v.)

The expected value and variance of the Poisson random variable $X \sim P(a)$ are

$$\boxed{\mu_X = a} \qquad \boxed{\sigma_X^2 = a}$$

Proof. Let's calculate the expected value making use of the definition

$$\mu_X = \sum_{k=0}^{\infty} k\, P_k = \sum_{k=0}^{\infty} k\, \frac{a^k}{k!} e^{-a} = a e^{-a} \sum_{k=1}^{\infty} \frac{a^{k-1}}{(k-1)!}$$

Letting $m = k - 1$, we have

$$\mu_X = a e^{-a} \sum_{m=0}^{\infty} \frac{a^m}{m!} = a e^{-a} e^{a} = a$$

□

Proof. For the variance

$$\sigma_X^2 = \sum_{k=0}^{\infty}(k-a)^2 \frac{a^k}{k!} e^{-a} = e^{-a} \sum_{k=0}^{\infty} k^2 \frac{a^k}{k!} + a^2 e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} - 2ae^{-a} \sum_{k=0}^{\infty} k \frac{a^k}{k!}$$
$$= \textbf{(I)} + \textbf{(II)} + \textbf{(III)}$$

(II) $a^2 e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} = a^2$

(III) $2ae^{-a} \sum_{k=0}^{\infty} k \frac{a^k}{k!} = 2ae^{-a}\left(ae^a\right) = 2a^2$

(I) $e^{-a} \sum_{k=0}^{\infty} k^2 \frac{a^k}{k!} = e^{-a} \sum_{k=1}^{\infty} k \frac{a^k}{(k-1)!} = ae^{-a} \sum_{k=1}^{\infty} (k-1+1) \frac{a^{k-1}}{(k-1)!} =$

$$ae^{-a} \left(\sum_{k=1}^{\infty} (k-1) \frac{a^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{a^{k-1}}{(k-1)!} \right) = ae^{-a}\left(ae^a + e^a\right) = a^2 + a$$

and finally: $\sigma_X^2 = a^2 + a + a^2 - 2a^2 = a$ $\square$

Example (1)

either 0 or 1

A noisy transmission channel has a per-digit error probability $p = 0.01$

1. Calculate the probability of more than one error in 10 received digits, using the binomial distribution
2. Repeat the calculation using the Poisson approximation

The number of errors in 10 received digits is a binomial random variable X with parameters $(n, p) = (10, 0.01)$. Then

$$\begin{aligned}\mathcal{P}(X > 1) &= 1 - \mathcal{P}(X = 0) - \mathcal{P}(X = 1) \\ &= 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 \\ &= 1 - 0.90438 - 0.09135 \simeq 0.00427\end{aligned}$$

Example (1):

Using the binomial dist. for more than one error in 10 received digits.

$$P(X>1) = 1 - P(X=0) - P(X=1)$$

$\hookrightarrow X \sim B(n,p)$

$X \sim B(20, \frac{1}{100})$

$$= 1 - \binom{10}{0}\left(\frac{1}{100}\right)^0\left(1-\frac{1}{100}\right)^{10} - \binom{10}{1}\left(\frac{1}{100}\right)^1\left(1-\frac{1}{100}\right)^9$$

$$\simeq 0.00427$$

Example (1)

A noisy transmission channel has a per-digit error probability $p = 0.01$

1. Calculate the probability of more than one error in 10 received digits, using the binomial distribution
2. Repeat the calculation using the Poisson approximation

Using of the Poisson approximation with $a = np = 0.01 \cdot 10 = 0.1$, we have

$$\begin{aligned}\mathcal{P}(X > 1) &= 1 - \mathcal{P}(X = 0) - \mathcal{P}(X = 1) \\ &= 1 - \frac{(0.1)^0}{0!}\, e^{-0.1} - \frac{(0.1)^1}{1!}\, e^{-0.1} \\ &= 1 - 0.90484 - 0.09048 \simeq 0.00468\end{aligned}$$

The error is $\sim 9.67\%$

Example (1):

Using Poisson approx. $\quad a = np = 10 \cdot 0.01 = 0.1$

$$P(X>1) = 1 - P(X=0) - P(X=1)$$
$$= 1 - \frac{(0.1)}{0!} e^{-0.1} - \frac{(0.1)^1 e^{-0.1}}{1!}$$
$$\simeq 0.00468$$

Example

What if the per-digit error probability $p = 0.001$ and there are $n = 100$ transmitted signals?

it does not make sense to use exact formula when the input data is already approximated

- Using the binomial distribution

$$\mathcal{P}(X > 1) = 1 - \mathcal{P}(X = 0) - \mathcal{P}(X = 1)$$
$$= 1 - \binom{100}{0}(0.001)^0 (0.999)^{100} - \binom{100}{1}(0.001)^1 (0.999)^{99}$$
$$= 1 - 0.9048 - 0.09057 \simeq 0.00464$$

- The Poisson approximation is the same as before ($a = np = 0.001 \cdot 100 = 0.1$)

$$\mathcal{P}(X > 1) = 1 - \frac{(0.1)^0}{0!} e^{-0.1} - \frac{(0.1)^1}{1!} e^{-0.1} \simeq 0.00468$$

The error is now $\sim 0.88\%$

Example (2)

The number of defective pieces assembled in any 10-hour period is known to be a Poisson random variable X with $a = 2$

1. Find the probability that more than 3 defective pieces will be assembled in a 10-hour period
2. Find the probability that no defective pieces will be assembled in 10-hour period

The probability mass function is $p_X(k) = \frac{2^k}{k!} e^{-2}$. In the first case

$$\mathcal{P}(X > 3) = 1 - \mathcal{P}(X \leqslant 3) = 1 - \sum_{k=0}^{3} \frac{2^k}{k!} e^{-2}$$

$$= 1 - e^{-2}\left(1 + 2 + \frac{4}{2} + \frac{8}{6}\right) \simeq 0.143$$

In the second case: $\mathcal{P}(X = 0) = e^{-2} \simeq 0.135$

Example (2)

Probability of >3 defects will be assembled in 10 hours.

$a = 2$, So to find $P(X>3) = 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3)$

$$P(X>3) = 1 - e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!}\right)$$

$$\simeq 0.143$$

So the prob. of none defect found is

$P(X=0) = e^{-2} \simeq 0.135$

Discrete Probability Distributions

trivial random variable

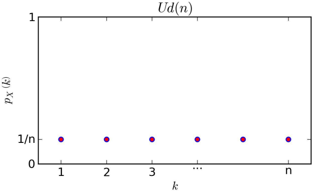
Definition (Discrete uniform random variable)

We say that X is a *discrete uniform random variable* if its probability mass function is constant

$$X \sim Ud(n)$$

If X can take on values from 1 to n, it must be:

$$\mathcal{P}(X = k) = \frac{1}{n} \quad k = 1, \ldots n$$



Theorem (Expected value and Variance of the discrete uniform r.v.)

Expected value and variance of the discrete uniform random variable are

$$\mu_X = \frac{n+1}{2} \qquad \sigma_X^2 = \frac{n^2-1}{12}$$