Discrete distributions

Four important discrete distributions:

1. The Uniform distribution (discrete)
2. The Binomial distribution
3. The Hyper-geometric distribution
4. The Poisson distribution

Uniform distribution
Definition

Experiment with *k* equally likely outcomes.

Definition:

Let X: S $\rightarrow$ R be a discrete random variable. If

$$P(X_1 = x_1) = P(X_2 = x_2) = \cdots P(X_k = x_k) = \frac{1}{k}$$

then the distribution of X is the (discrete) uniform distribution.

Probability function:

$$f(x:k) = \frac{1}{k} \quad for \;\; x = x_1, x_2, \ldots, x_k$$

(Cumulative) distribution function:

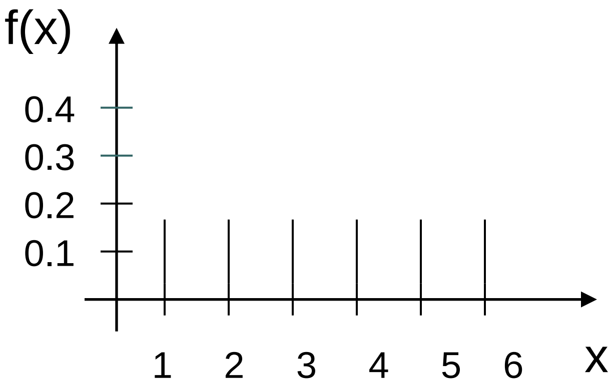
$$F(x;k) = \frac{x}{k} \quad for \;\; x = x_1, x_2, \ldots, x_k$$

Uniform distribution

Example

Example: Rolling a dice

X: # eyes



Mean value:

$$E(X) = \frac{1+2+3+4+5+6}{6} = 3.5$$

variance:

$$Var(X) = \frac{(1-3.5)^2 + \ldots + (6-3.5)^2}{6}$$

$$= \frac{35}{12}$$

Probability function:

$$f(x;k) = \frac{1}{6} \quad for\ x = 1,2,\ldots,6$$

Distribution function:

$$F(x;6) = \frac{x}{6} \quad for\ x = 1,2,\ldots,6$$

Uniform distribution
Mean & variance

Theorem:
Let X be a uniformly distributed with outcomes $x_1, x_2, \ldots, x_k$
Then we have

- mean value of X: $$E(X) = \mu = \frac{\sum_{i=1}^{k} x_i}{k}$$
- variance af X: $$Var(X) = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}$$

Binomial distribution
Bernoulli process

Repeating an experiment with *two* possible outcomes.

Bernoulli process:

1. The experiment consists in repeating the same trial n times.
2. Each trial has two possible outcomes: "*success*" or "*failure*", also known as Bernoulli trial.
3. P("succes") = p is the same for all trials.
4. The trials are independent.

Binomial distribution
Bernoulli process

Definition:
Let the random variable X be the number of "successes" in the n Bernoulli trials.

The distribution of X is called the binomial distribution**.**
Notation: $X \sim \text{bin}(n,p)$

Binomial distribution
Probability & distribution function

Theorem:
If X ~ bin(n,p), then X has probability function

$$b(x\,;n,p)=P(X=x)=\binom{n}{x}p^x(1-p)^{n-x},\quad x=0,1,2,\ldots,n$$

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

and distribution function

$$B(x\,;n,p)=P(X\le x)=\sum_{t=0}^{x}b(t;n,p),\quad x=0,1,2,\ldots,n$$ (See Table A.1)

Binomial distribution
Problem

LEGO has a policy of discarding a batch of bricks if they do not fulfil LEGO's "quality control":

• A sample of 20 LEGO bricks is taken: If one or more bricks are defective, the entire batch is discarded.
• Assume the batch contains 10% defective bricks.

1. What is the probability that the entire batch is discarded?

2. What is the probability that at most 3 bricks are defective?

Binomial distribution
Mean & variance

Theorem:
If X ~ bin(n,p), then

- mean of X: $E(X) = np$
- variance of X: $Var(X) = np(1-p)$

Example continued:
What is the expected number of defective bricks?

Binomial distribution
Tables and MATLAB

Appendix A of [W], Table A.1

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x;n,p)$

		p							
n	*r*	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
20	**0**	0.1216	0.0115	0.0032	0.0008	0.0000			
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000		
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002		
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0000	
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.000

In MATLAB:

`binopdf(X,N,P)=` $\binom{N}{X} P^{X} (1-P)^{N-X}$

`binocdf(X,N,P)=` $\sum_{x=0}^{X} \binom{N}{x} P^{x} (1-P)^{N-x}$

Hyper-geometric distribution
Hyper-geometric experiment

Hyper-geometric experiment:

1. n elements chosen from N elements without replacement.
2. k of these N elements are ”successes” and N-k are ”failures”

Notice!! Unlike the binomial distribution the selection is done without replacement and the experiments and not independent.

Often used in quality control.

Hyper-geometric distribution
Definition

Definition:
Let the random variable X be the number of "successes" in a hyper-geometric experiment, where n elements are chosen from N elements, of which k are "successes" and N-k are "failures".

The distribution of X is called the hyper-geometric distribution.

Notation: X ~ hypergeo(N, n, k)

Hyper-geometric distribution
Probability & distribution function

Theorem:

If X ~ hypergeo(N, n, k), then X has probability function

$$h(x; N, n, k) = P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, n$$

and distribution function

$$H(x; N, n, k) = P(X \leq x) = \sum_{t=0}^{x} h(t; N, n, k), \quad x = 0, 1, 2, \ldots, n$$

Hyper-geometric distribution
Problem

Toy 'R' Us receives a shipment of 40 LEGO bricks. The shipment is unacceptable if 3 or more bricks are defective.

Sample plan: take 5 bricks. If at least one brick is defective the entire shipment is rejected.

What is the probability of exactly one defective brick, if the shipment contains 3 defective bricks ?

Is this a good sample plan ?

Hyper-geometric distribution
Mean & variance

Theorem:
If X ~ hypergeo(N, n, k), then

- mean of X: $$E(X) = \frac{n\,k}{N}$$
- variance of X: $$Var(X) = \frac{N-n}{N-1}\, n \,\frac{k}{N}\left(1 - \frac{k}{N}\right)$$

Hyper-geometric distribution
MATLAB

There are no tables in [W]

In MATLAB:

`hygepdf(X,M,K,N)` $= \dfrac{\dbinom{K}{X}\dbinom{M-K}{N-X}}{\dbinom{M}{N}}$

`hygecdf(X,M,K,N)` $= \displaystyle\sum_{x=0}^{X} \dfrac{\dbinom{K}{x}\dbinom{M-K}{N-x}}{\dbinom{M}{N}}$

Poisson distribution
Poisson process

Experiment where events are observed during a time interval.

Poisson process:

1. # events in the interval [a,b] is independent of # events in the interval [c,d], where a<b<c<d } **No memory**
2. Probability of 1 event in a short time interval [a, a + ε] is proportional to ε.
3. The probability of more then 1 event in the short time interval is close to 0.

Poisson distribution
Definition

Definition:
Let the random variable X be the number of events in a time interval of length t from a Poisson process, which has on average λ events pr. unit time.

The distribution of X is called the Poisson distribution with parameter $\mu = \lambda t$.

Notation: $X \sim \text{Pois}(\mu)$, where $\mu = \lambda t$

Poisson distribution
Probability & distribution function

Theorem:

If $X \sim \mathrm{Pois}(\mu)$, then X has probability function

$$p(x\,;\mu) = P(X = x) = \frac{e^{-\mu}\mu^{x}}{x!}, \quad x = 0, 1, 2, \ldots$$

and distribution function

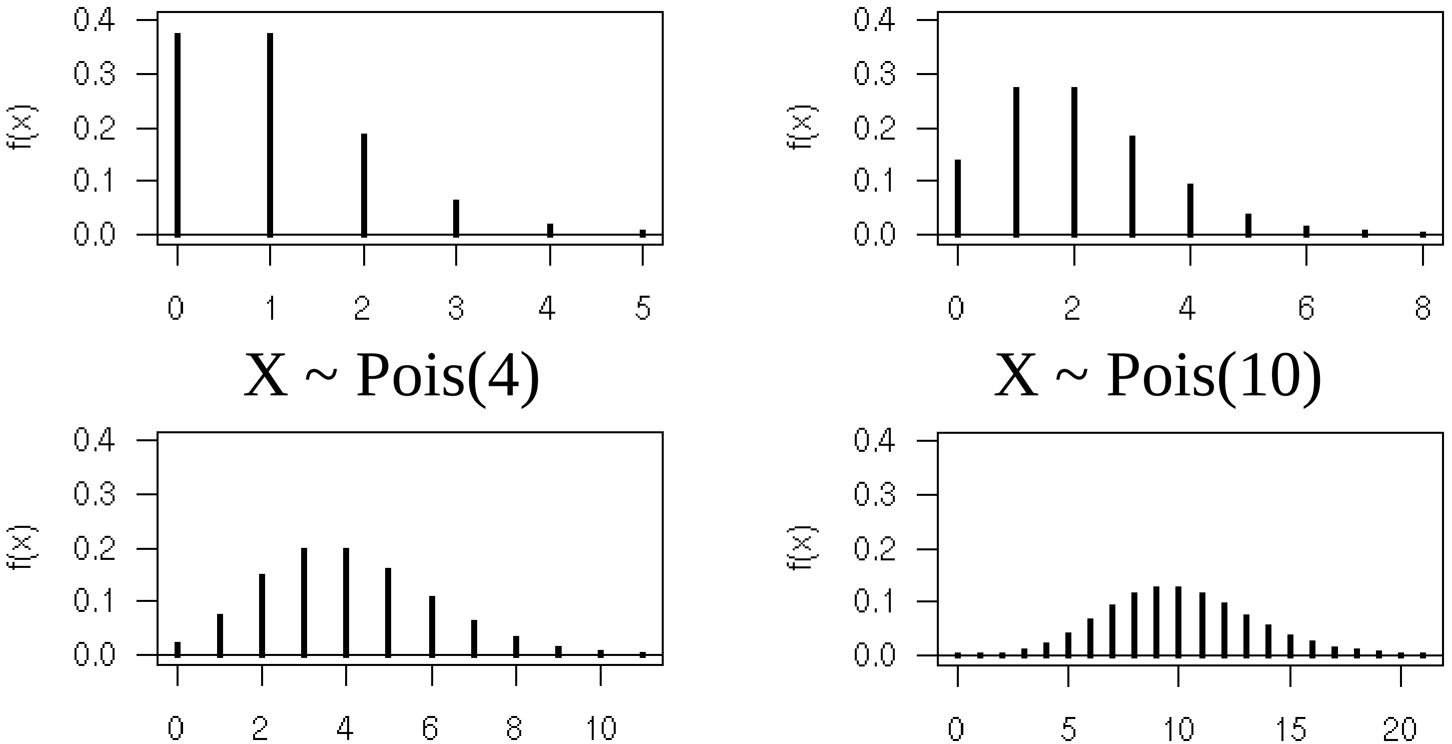
$$P(x\,;\mu) = P(X \le x) = \sum_{t=0}^{x} p(t\,;\mu), \quad x = 0, 1, 2, \ldots$$ (see Table A2)

Poisson distribution
Examples

Some examples of $X \sim \text{Pois}(\mu)$:

$X \sim \text{Pois}(1)$ $X \sim \text{Pois}(2)$

$X \sim \text{Pois}(4)$ $X \sim \text{Pois}(10)$



Poisson distribution
Mean & variance

Theorem:
If $X \sim \mathrm{Pois}(\mu)$, then

- mean of X: $E(X) = \mu$
- variance of X: $\mathrm{Var}(X) = \mu$

Poisson distribution
Tables and MATLAB

Appendix A of [W], Table A.2

Table A.2 Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

	μ						
r	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
0	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.018
1	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.091
2	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.238
3	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.433
4	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.628
5	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.785
6	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.889

In MATLAB:

`poisspdf(X,lambda)` $= \dfrac{e^{-\lambda}\lambda^{X}}{X!}$

`poisscdf(X,lambda)` $= \sum_{x=0}^{X} \dfrac{e^{-\lambda}\lambda^{x}}{x!}$

Poisson distribution
Problem

Politiken.dk has done some research: On weekdays before noon an average of 3 clients enters their website per minute.

1. What is the probability that exactly 2 clients enter during the time interval 11.38 - 11.39 (i.e. one minute)
2. What is the probability that at least 2 clients enter in the interval above?
3. What is the probability that at least 10 clients enter the site in the time interval 10.05 - 10.10?