Probability and Statistics

Kristel Van Steen, PhD[2]

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@ulg.ac.be

CHAPTER 3: PARAMETRIC FAMILIES OF UNIVARIATE DISTRIBUTIONS

1 Why do we need distributions?

1.1 Some practical uses of probability distributions

1.2 Related distributions

1.3 Families of probability distributions

2 Discrete distributions

2.1 Introduction

2.2 Discrete uniform distributions

2.3 Bernoulli and binomial distribution

2.4 Hypergeometric distribution

2.5 Poisson distribution

3 Continuous distributions

3.1 Introduction

3.2 Uniform or rectangular distribution

3.3 Normal distribution

3.4 Exponential and gamma distribution

3.5 Beta distribution

4 Where discrete and continuous distributions meet

4.1 Approximations

4.2 Poisson and exponential relationships

4.3 Deviations from the ideal world ?

4.3.1 Mixtures of distributions

4.3.2 Truncated distributions

5 Conditional distributions and stochastic independence

5.1 Conditional distribution functions for discrete random variables

5.2 Conditional distribution functions for continuous random variables

1 Why do we need distributions?

Probability distributions are a fundamental concept in statistics. They are used both on a theoretical level and a practical level.

1.1 Some practical uses of probability distributions

- To calculate confidence intervals for parameters and to calculate critical regions for hypothesis tests.
- For univariate data, it is often useful to determine a reasonable distributional model for the data.

- Statistical intervals and hypothesis tests are often based on specific distributional assumptions. Before computing an interval or test based on a distributional assumption, we need to verify that the assumption is justified for the given data set. In this case, the distribution does not need to be the best-fitting distribution for the data, but an adequate enough model so that the statistical technique yields valid conclusions.
- Simulation studies with random numbers generated from using a specific probability distribution are often needed.

Recall

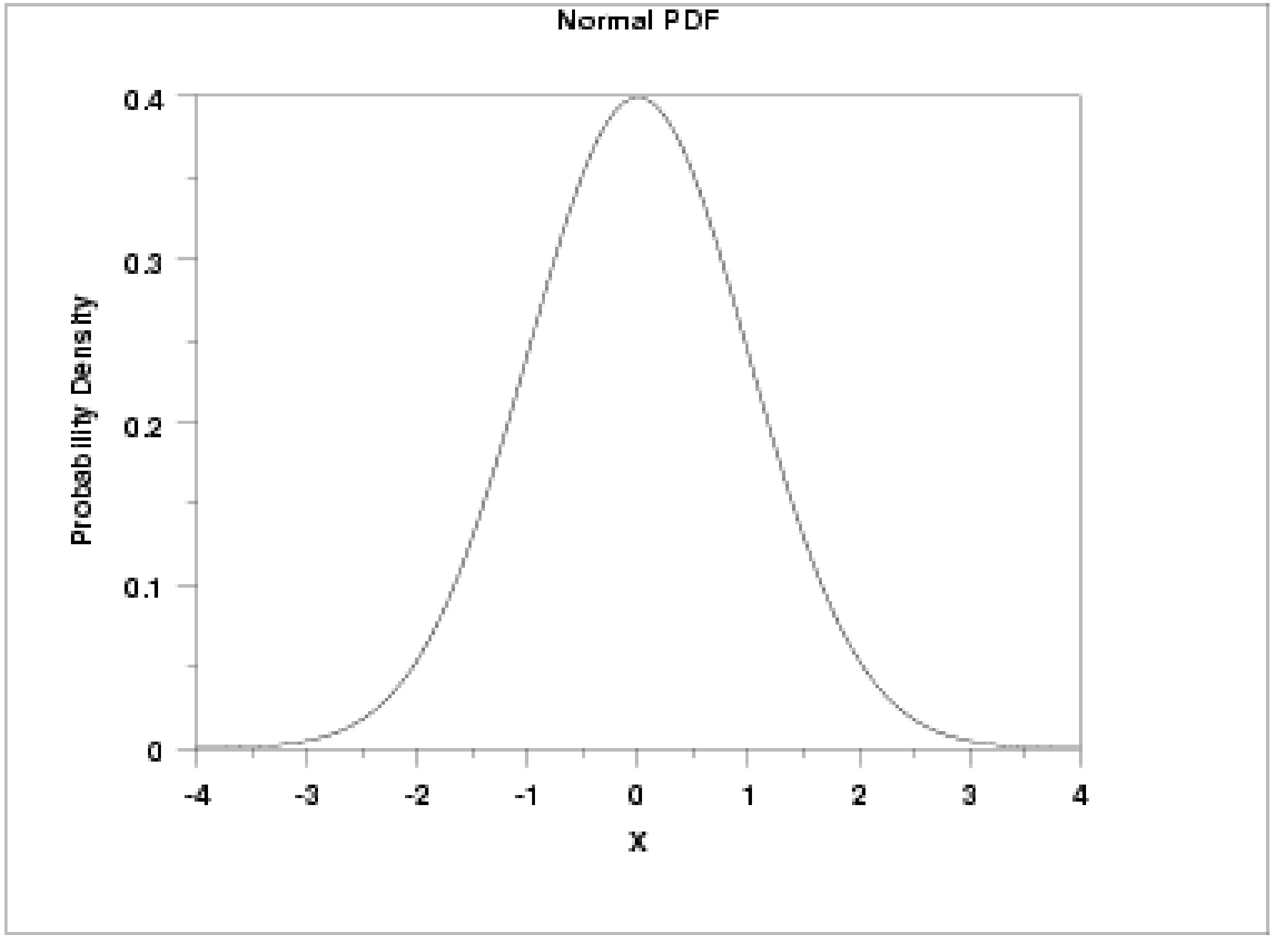
- For a continuous function, the probability density function (pdf) is the probability that the variate has the value *x*. Since for continuous distributions the probability at a single point is zero, this is often expressed in terms of an integral between two points.

$$\int_a^b f(x)dx = Pr[a \leq X \leq b]$$

- For a discrete distribution, the pdf is the probability that the variate takes the value x.

$$f(x) = Pr[X = x]$$

- The following is the plot of the normal probability density function.



1.2 Related distributions

- The **cumulative distribution function** (cdf) is the probability that the variable takes a value less than or equal to x. That is

$$F(x) = Pr[X \leq x] = \alpha$$

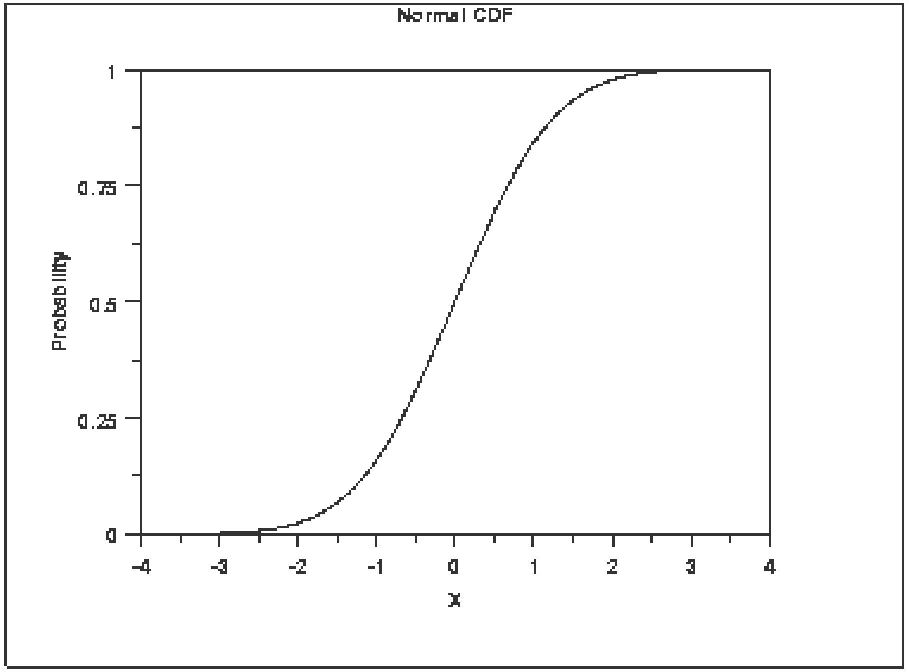
  - For a continuous distribution, this can be expressed mathematically as

$$F(x) = \int_{-\infty}^{x} f(\mu) d\mu$$

  - For a discrete distribution, the cdf can be expressed as

$$F(x) = \sum_{i=0}^{x} f(i)$$

- The following is the plot of the normal cumulative distribution function.



- The horizontal axis is the allowable domain for the given probability function. Since the vertical axis is a probability, it must fall between zero and one. It increases from zero to one as we go from left to right on the horizontal axis.

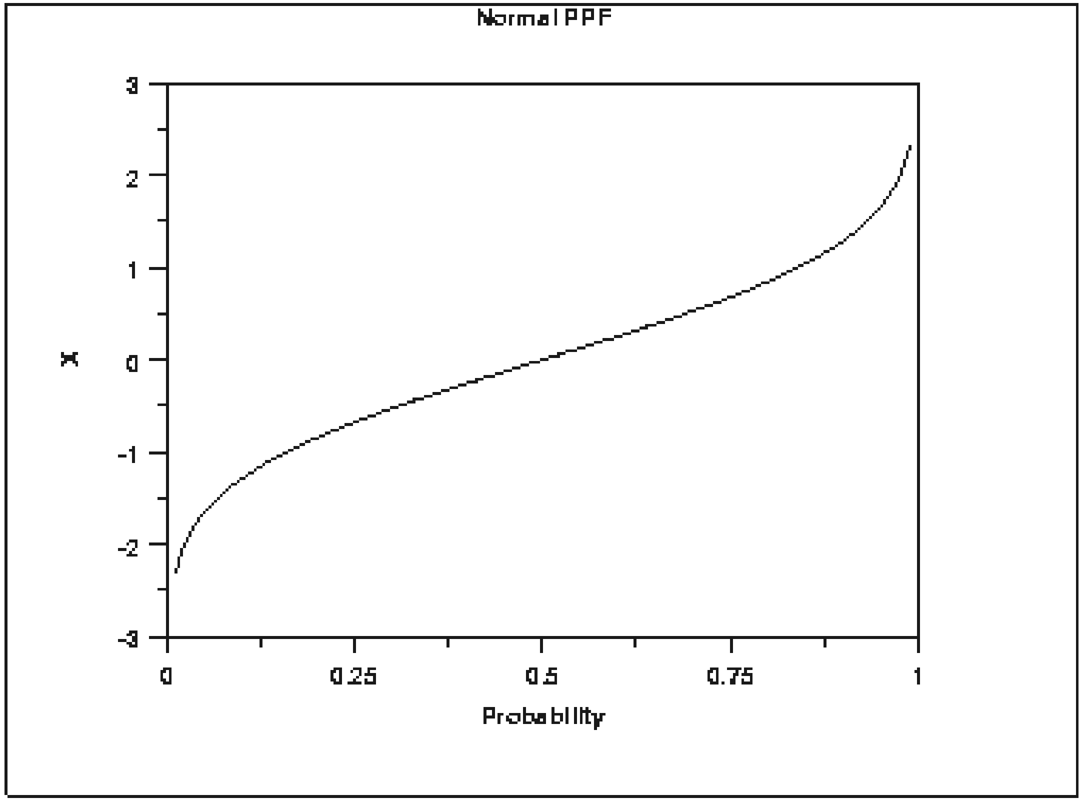
- The **percent point function** (ppf) is the inverse of the cumulative distribution function.
- For this reason, the percent point function is also commonly referred to as the **inverse distribution function**.
  - o That is, for a distribution function we calculate the probability that the variable is less than or equal to *x* for a given *x*.
  - o For the percent point function, we start with the probability and compute the corresponding *x* for the cumulative distribution.
- Mathematically, this can be expressed as

$$Pr[X \leq G(\alpha)] = \alpha$$

or alternatively

$$x = G(\alpha) = G(F(x))$$

- The following is the plot of the normal percent point function.

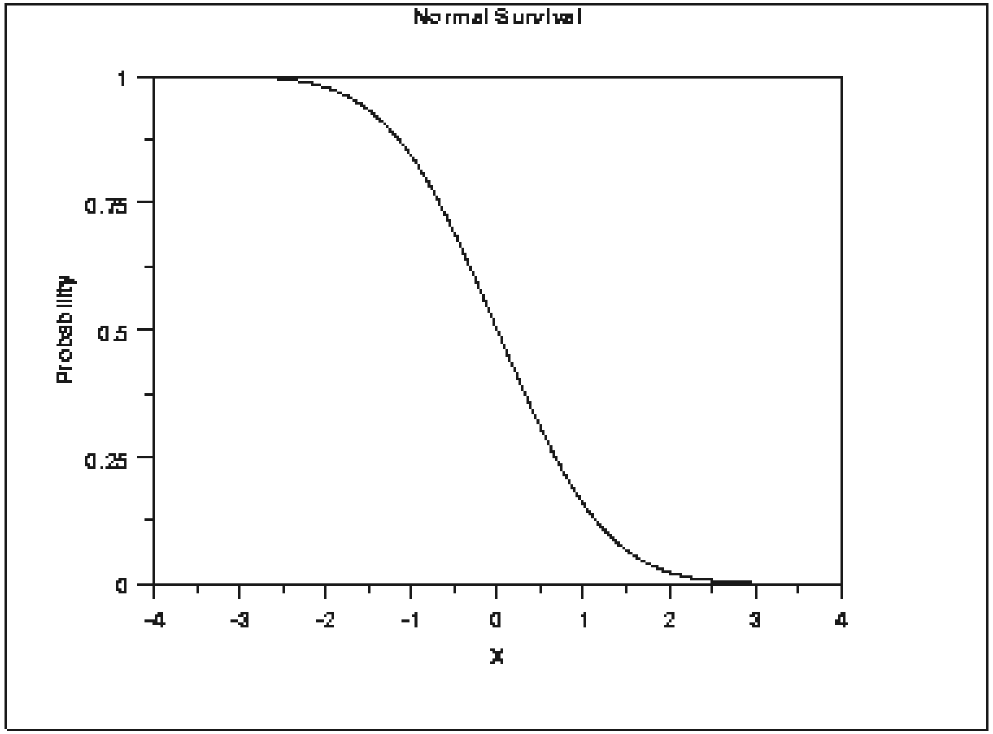


- Since the horizontal axis is a probability, it goes from zero to one. The vertical axis goes from the smallest to the largest value of the cumulative distribution function.

- Survival functions are most often used in reliability and related fields. The survival function is the probability that the variate takes a value greater than *x*.

$$S(x) = Pr[X > x] = 1 - F(x)$$

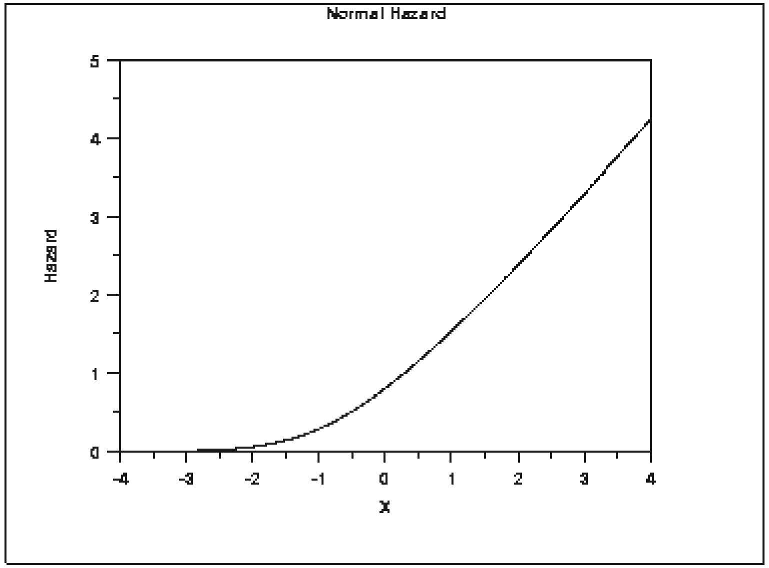
- The following is the plot of the normal distribution survival function.



- For a survival function, the *y* value on the graph starts at 1 and monotonically decreases to zero.
- The survival function should be compared to the cumulative distribution function.
- The **hazard function** is the ratio of the probability density function to the survival function, *S*(x).

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)}$$

- The following is the plot of the normal distribution hazard function.



- Hazard plots are most commonly used in reliability applications (sometimes referred to as conditional failure density function).

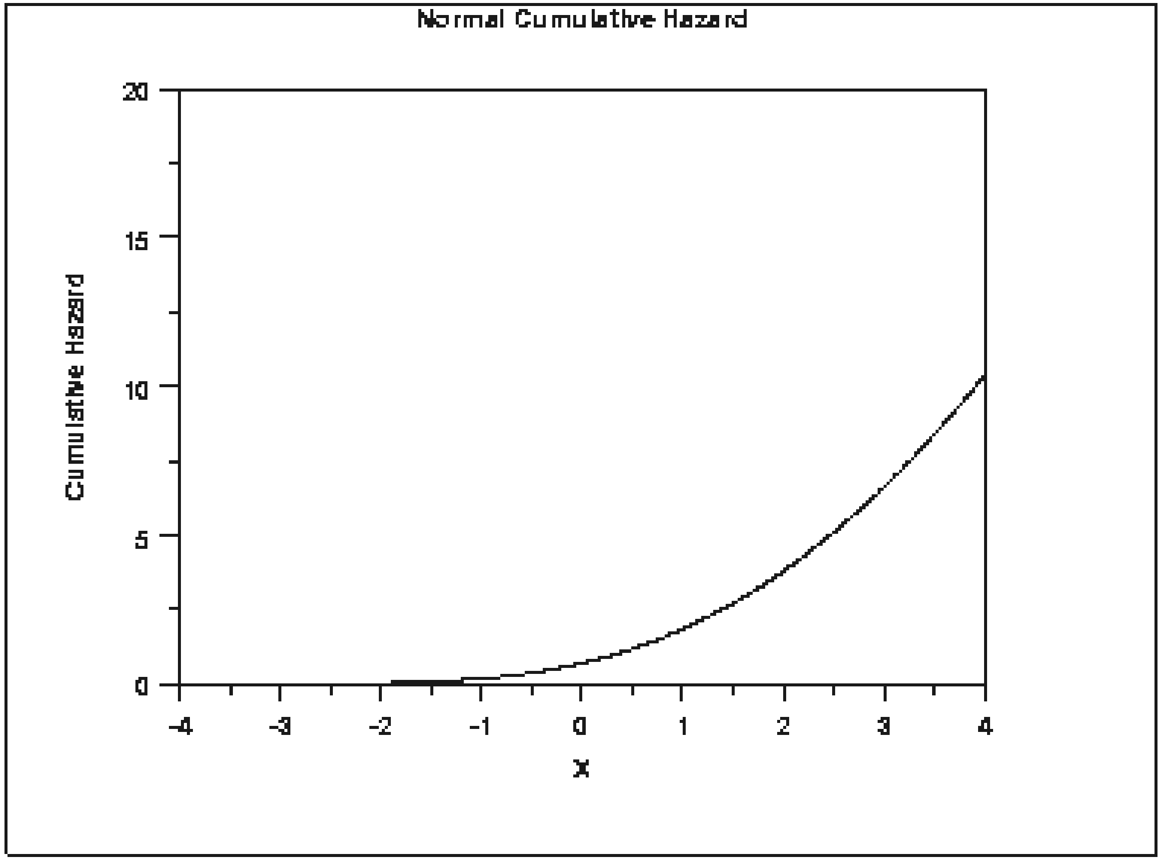
- The **cumulative hazard function** is the integral of the hazard function. It can be interpreted as the probability of failure at time x given survival until time *x*.

$$H(x) = \int_{-\infty}^{x} h(\mu) d\mu$$

- This can alternatively be expressed as

$$H(x) = -\ln(1 - F(x))$$

- The following is the plot of the normal cumulative hazard function.



- Cumulative hazard plots are most commonly used in reliability applications.

1.3 Families of distributions

- Many probability distributions are not a single distribution, but are in fact a family of distributions. This is due to the distribution having one or more shape parameters.
- Shape parameters allow a distribution to take on a variety of shapes, depending on the value of the shape parameter.
- These distributions are particularly useful in modeling applications since they are flexible enough to model a variety of data sets.

Example: the Weibull distribution

There are many probability distributions beyond the binomial and normal distributions used to model data in various circumstances.

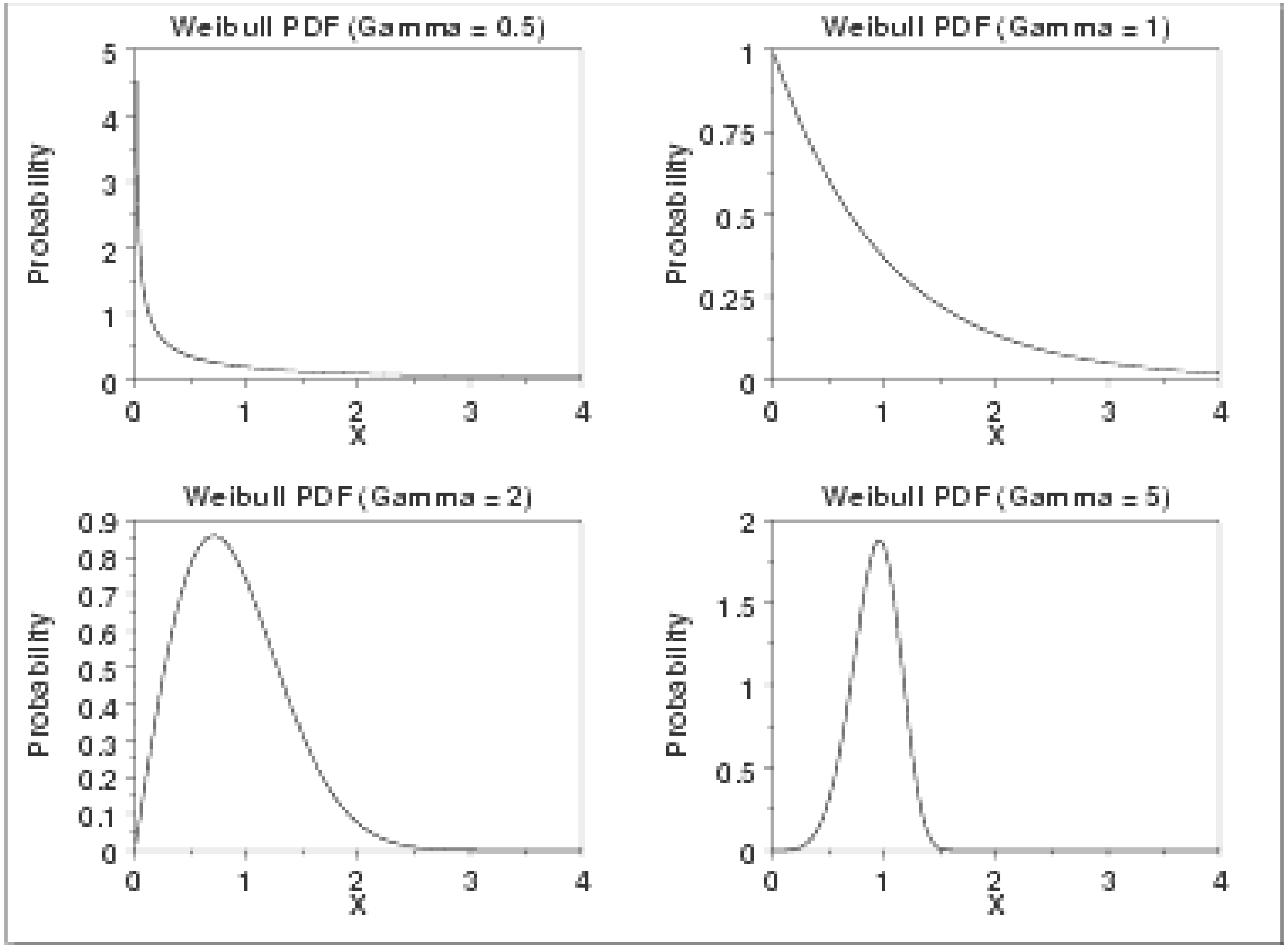
Weibull distributions are used to model **time to failure/product lifetime** and are common in engineering to study product reliability.

Product lifetimes can be measured in units of time, distances, or number of cycles for example. Some applications include:

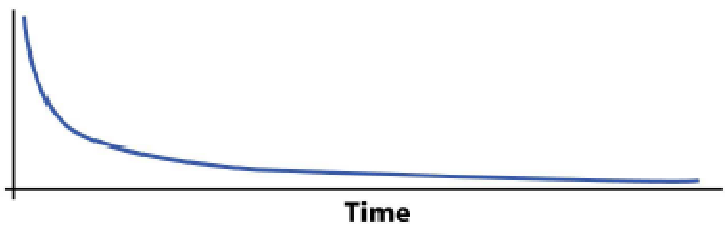
- Quality control (breaking strength of products and parts, food shelf life)
- Maintenance planning (scheduled car revision, airplane maintenance)
- Cost analysis and control (number of returns under warranty, delivery time)
- Research (materials properties, microbial resistance to treatment)

- The Weibull distribution is an example of a distribution that has a shape parameter.
- The shapes on the next slide include an exponential distribution, a right-skewed distribution, and a relatively symmetric distribution.
- So although the Weibull distribution has a relatively simple distributional form (see later), the shape parameter allows the Weibull to assume a wide variety of shapes.
- This combination of simplicity and flexibility in the shape of the Weibull distribution has made it an effective distributional model in reliability applications.
- This ability to model a wide variety of distributional shapes using a relatively simple distributional form is possible with many other distributional families as well.

- The following graph plots the Weibull pdf with the following values for the shape parameter: 0.5, 1.0, 2.0, and 5.0.

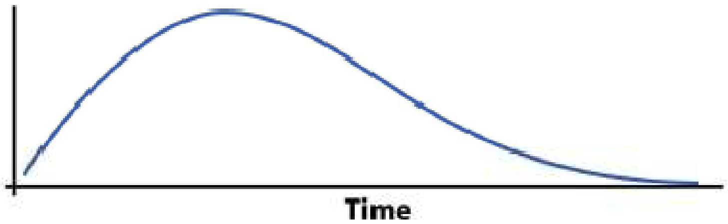


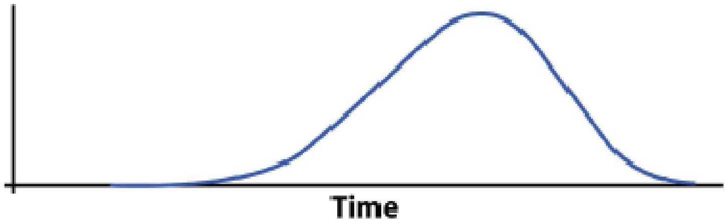
Density curves of three members of the Weibull family describing a different type of product time to failure in manufacturing:



Infant mortality: Many products fail immediately and the remainders last a long time. Manufacturers only ship the products after inspection.

Early failure: Products usually fail shortly after they are sold. The design or production must be fixed.





Old-age wear out: Most products wear out over time, and many fail at about the same age. This should be disclosed to customers.

The standard form of a distribution

Definition

The **standard form of any distribution** is the form that has location parameter zero and scale parameter one.

- It is common in statistical software packages to only compute the standard form of the distribution.
- There are formulas for converting from the standard form to the form with other location and scale parameters.
- These formulas are independent of the particular probability distribution.

- The following are the formulas for computing various probability functions based on the standard form of the distribution. In what follows, the parameter *a* refers to the **location parameter** and the parameter *b* refers to the **scale parameter**. Shape parameters are not included.

Cumulative Distribution Function	$F(x;a,b) = F((x-a)/b;0,1)$
Probability Density Function	$f(x;a,b) = (1/b)f((x-a)/b;0,1)$
Percent Point Function	$G(\alpha;a,b) = a + bG(\alpha;0,1)$
Hazard Function	$h(x;a,b) = (1/b)h((x-a)/b;0,1)$
Cumulative Hazard Function	$H(x;a,b) = H((x-a)/b;0,1)$
Survival Function	$S(x;a,b) = S((x-a)/b;0,1)$
Random Numbers	$Y(a,b) = a + bY(0,1)$

Note

- A **location parameter** simply shifts the graph left (location parameter is negative) or right (location parameter is positive) on the horizontal axis
- The effect of a **scale parameter** greater than one is to stretch the pdf. The greater the magnitude, the greater the stretching. The effect of a scale parameter less than one is to compress the pdf. The compressing approaches a spike as the scale parameter goes to zero.
- A third characteristic of a distribution is its **shape**. The shape shows how the variation is distributed about the location. This tells us if our variation is symmetric about the mean or if it is skewed or possibly multimodal.

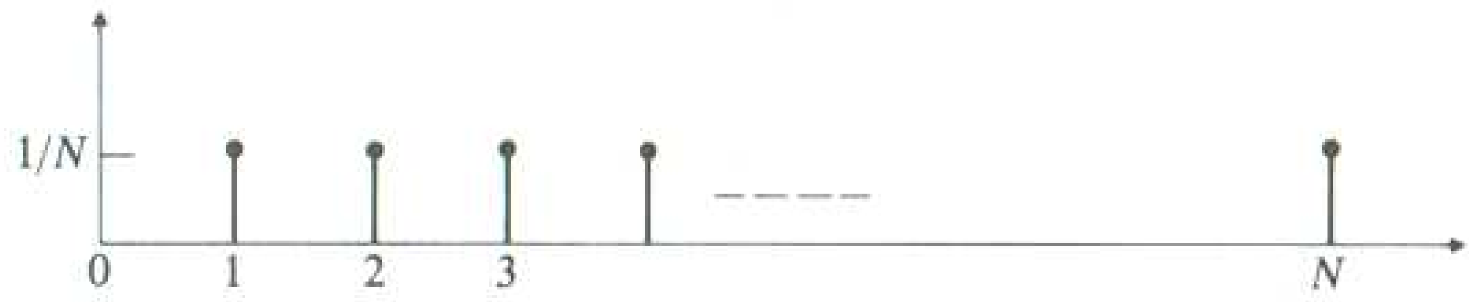
2 Discrete distributions

2.1 Introduction

Distribution	Probability Mass Function $p(x)$	Mean	Variance	Moment Generating Function
Binomial binomial(n,p)	$\binom{n}{x} p^x q^{n-x}, x = 0, 1, \cdots, n$	np	npq	$(pe^t + q)^n$
Geometric $G(p)$	$(i)\ pq^x, x = 0, 1, \cdots$ $(ii)\ pq^{y-1}, y = 1, 2, \cdots$	$(i)\ q/p$ $(ii)\ 1/p$	$(i)\ q/p^2$ $(ii)\ q/p^2$	$(i)\ p/(1-qe^t)$ $(ii)\ pe^t/(1-qe^t)$
Hypergeometric $H(n,a,N)$	$\binom{a}{x}\binom{N-a}{n-x} / \binom{N}{n}$ $x = 0, 1, 2, \cdots, \min(N-a, n)$	np $p = a/N$	$\frac{(N-n)}{(N-1)} npq$	complicated

Distribution	Probability Mass Function $p(x)$	Mean	Variance	Moment Generating Function
Poisson Poisson(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, \cdots$	λ	λ	$e^{\lambda(e^t-1)}$
Negative Binomial	$(i)\ \binom{x+r-1}{x} p^r q^x, x = 0, 1, \cdots$	$(i)\ rq/p$	$(i)\ rq/p^2$	$(i)\ [p/(1-qe^t)]^r$
$NB(r,p)$	$(ii)\ \binom{x-1}{r-1} p^r q^{x-r}, x = r, r+1, \cdots$	$(ii)\ r/p$	$(ii)\ rq/p^2$	$(ii)\ [pe^t/(1-qe^t)]^r$

2.2 Discrete uniform distributions



Definition **Discrete uniform distribution** Each member of the family of discrete density functions

$$f(x) = f(x; N) = \begin{cases} \frac{1}{N} & \text{for } x = 1, 2, \ldots, N \\ 0 & \text{otherwise} \end{cases} = \frac{1}{N} I_{\{1, 2, \ldots, N\}}(x),$$

where the parameter N ranges over the positive integers, is defined to have a *discrete uniform distribution*. A random variable X having a density given in Eq. is called a *discrete uniform random variable*. ////

Theorem If X has a discrete uniform distribution, then $\mathscr{E}[X] = (N+1)/2$,

$$\text{var}\,[X] = \frac{(N^2-1)}{12}, \text{ and } m_X(t) = \mathscr{E}[e^{tX}] = \sum_{j=1}^{N} e^{jt}\,\frac{1}{N}.$$

Proof

$$\mathscr{E}[X] = \sum_{j=1}^{N} j \frac{1}{N} = \frac{N+1}{2}.$$

$$\operatorname{var}[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = \sum_{j=1}^{N} j^2 \frac{1}{N} - \left(\frac{N+1}{2}\right)^2$$

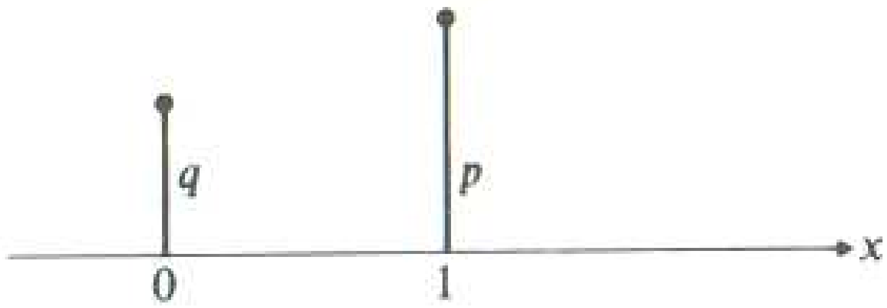
$$= \frac{N(N+1)(2N+1)}{6N} - \frac{(N+1)^2}{4} = \frac{(N+1)(N-1)}{12}.$$

$$\mathscr{E}[e^{tX}] = \sum_{j=1}^{N} e^{jt} \frac{1}{N}.$$

////

2.3 Bernoulli and binomial distribution

Bernoulli density



Definition **Bernoulli distribution** A random variable X is defined to have a *Bernoulli distribution* if the discrete density function of X is given by

$$f_X(x) = f_X(x; p)$$
$$= \begin{cases} p^x(1-p)^{1-x} & \text{for } x = 0 \text{ or } 1 \\ 0 & \text{otherwise} \end{cases} = p^x(1-p)^{1-x} I_{\{0,1\}}(x),$$

where the parameter p satisfies $0 \leq p \leq 1$. $1 - p$ is often denoted by q. ////

Theorem If X has a Bernoulli distribution, then

$$\mathscr{E}[X] = p, \qquad \text{var}[X] = pq, \qquad \text{and} \qquad m_X(t) = pe^t + q.$$

PROOF $\mathscr{E}[X] = 0 \cdot q + 1 \cdot p = p.$

$$\text{var}[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = 0^2 \cdot q + 1^2 \cdot p - p^2 = pq.$$

$$m_X(t) = \mathscr{E}[e^{tX}] = q + pe^t.$$

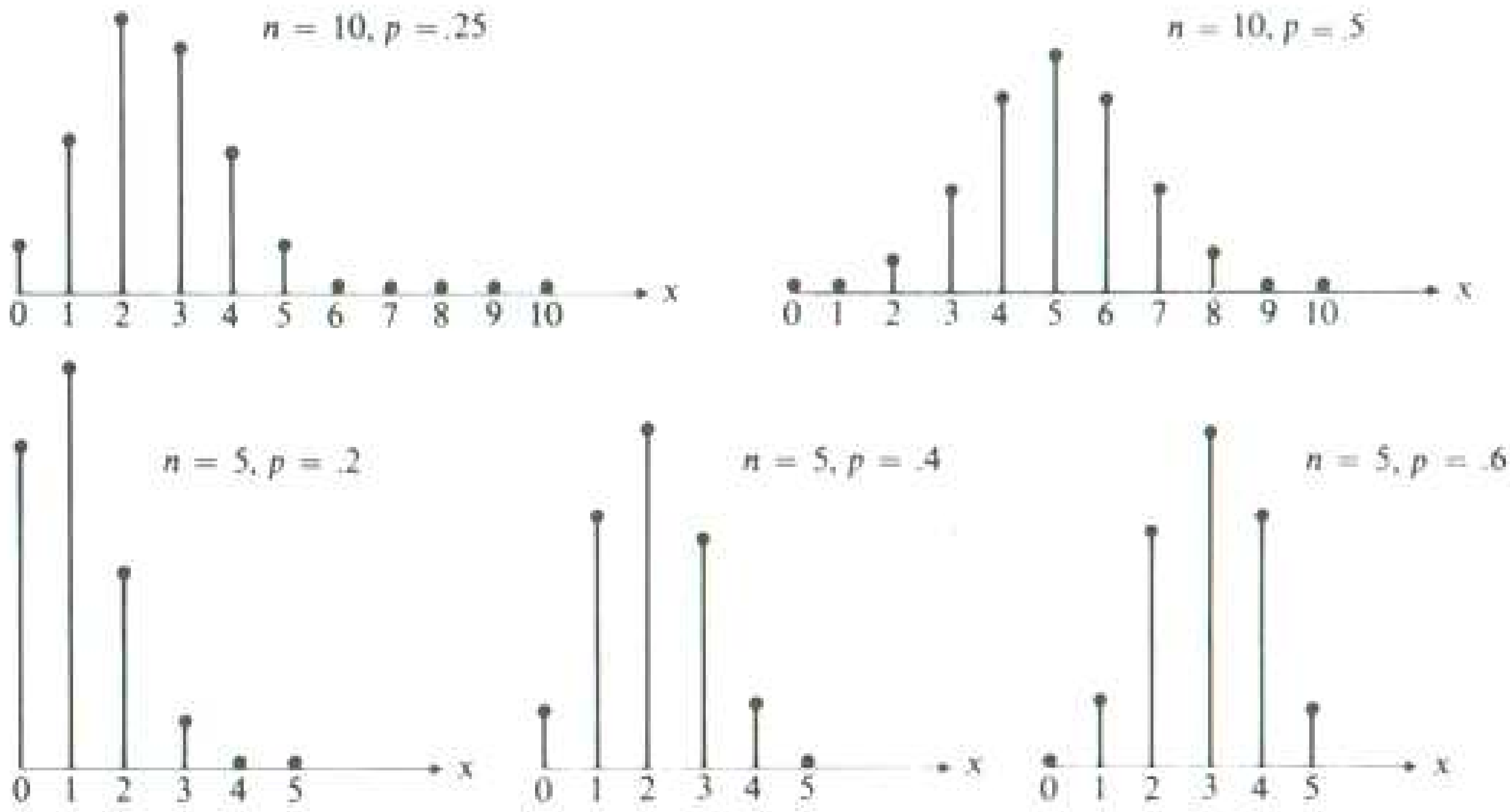
////

Examples

EXAMPLE 1 A random experiment whose outcomes have been classified into two categories, called "success" and "failure," represented by the letters $\mathscr{s}$ and $\mathscr{f}$, respectively, is called a *Bernoulli trial*. If a random variable X is defined as 1 if a Bernoulli trial results in success and 0 if the same Bernoulli trial results in failure, then X has a Bernoulli distribution with parameter $p = P[\text{success}]$. ////

EXAMPLE 2 For a given arbitrary probability space $(\Omega, \mathscr{A}, P[\cdot])$ and for A belonging to $\mathscr{A}$, define the random variable X to be the indicator function of A; that is, $X(\omega) = I_A(\omega)$; then X has a Bernoulli distribution with parameter $p = P[X = 1] = P[A]$. ////

Binomial distribution



Definition **Binomial distribution** A random variable X is defined to have a *binomial distribution* if the discrete density function of X is given by

$$f_X(x) = f_X(x; n, p) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{for } x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

$$= \binom{n}{x} p^x q^{n-x} I_{\{0, 1, \ldots, n\}}(x),$$

Theorem If X has a binomial distribution, then

$$\mathscr{E}[X] = np, \qquad \text{var}\,[X] = npq, \qquad \text{and} \qquad m_X(t) = (q + pe^t)^n.$$

Proof

$$m_X(t) = \mathscr{E}[e^{tX}] = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x q^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x q^{n-x}$$
$$= (pe^t + q)^n.$$

Now

$$m'_X(t) = npe^t(pe^t + q)^{n-1}$$

and

$$m''_X(t) = n(n-1)(pe^t)^2(pe^t + q)^{n-2} + npe^t(pe^t + q)^{n-1};$$

hence

$$\mathscr{E}[X] = m'_X(0) = np$$

and

$$\text{var}[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2$$
$$= m''_X(0) - (np)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p). \qquad ////$$

Common statistics

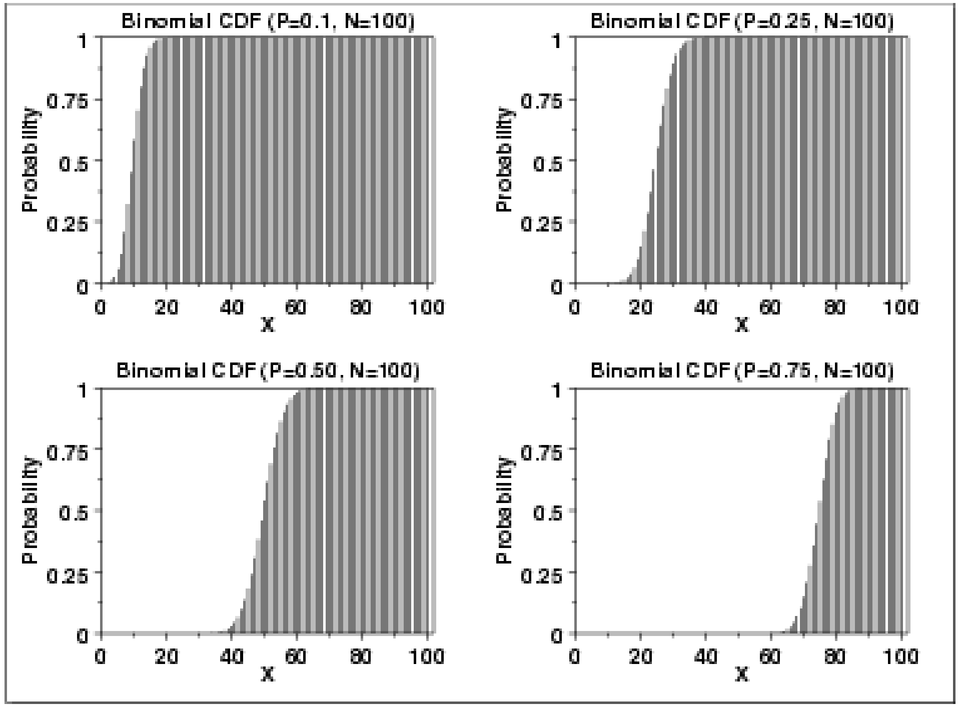
Mean	np
Mode	$p(n+1)-1 \leq x \leq p(n+1)$
Range	0 to N
Standard Deviation	$\sqrt{np(1-p)}$
Coefficient of Variation	$\sqrt{\frac{(1-p)}{np}}$
Skewness	$\frac{(1-2p)}{\sqrt{np(1-p)}}$
Kurtosis	$3-\frac{6}{n}+\frac{1}{np(1-p)}$

Cumulative distribution function

- The formula for the binomial cumulative probability function is

$$F(x,p,n) = \sum_{i=0}^{x} \binom{n}{i} (p)^i (1-p)^{(n-i)}$$

- The following is the plot of the binomial cumulative distribution function.



Example

- The binomial distribution is used when there are exactly two mutually exclusive outcomes of a trial.
- These outcomes are appropriately labeled "success" and "failure".
- The binomial distribution is used to obtain the probability of observing x successes in N trials, with the probability of success on a single trial denoted by p.

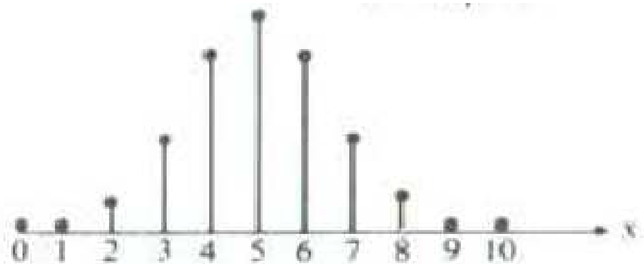
- In a clinical trial, a patient's condition may improve or not. We study the number of patients who improved, not how much better they feel.
- Is a person ambitious or not? The binomial distribution describes the number of ambitious persons, not how ambitious they are.
- In quality control we assess the number of defective items in a lot of goods, irrespective of the type of defect.

Consider sampling with replacement from an urn containing M balls, K of which are defective. Let X represent the number of defective balls in a sample of size n. The individual draws are Bernoulli trials where "defective" corresponds to "success," and the experiment of taking a sample of size n with replacement consists of n repeated independent Bernoulli trials where $p = P[\text{success}] = K/M$; so X has the binomial distribution

$$\binom{n}{x}\left[\frac{K}{M}\right]^x\left[1-\frac{K}{M}\right]^{n-x} \qquad \text{for} \quad x = 0, 1, \ldots, n,$$

Furthermore

- So the binomial distribution assumes that *p* is fixed for all trials.
- The binomial distribution reduces to the Bernoulli distribution when n=1. Therefore, sometimes the Bernoulli distribution is called the point binomial distribution
- From the graphical representations it is clear that the binomial distribution first increases monotonically and then decreases monotonically



Binomial formulas

The number of ways of arranging *k* successes in a series of *n* observations (with constant probability *p* of success) is the number of possible combinations (unordered sequences).

This can be calculated with the **binomial coefficient**:

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

Where k = 0, 1, 2, ..., or n.

- The binomial coefficient "*n*_choose_*k*" uses the **factorial** notation "**!**".
- The factorial $n!$ for any strictly positive whole number n is:

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1$$

- For example: $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$
- Note that $0! = 1$.

The binomial coefficient counts the number of ways in which k successes can be arranged among n observations.

The **binomial probability $P(X = k)$** is this count multiplied by the probability of any specific arrangement of the *k* successes:

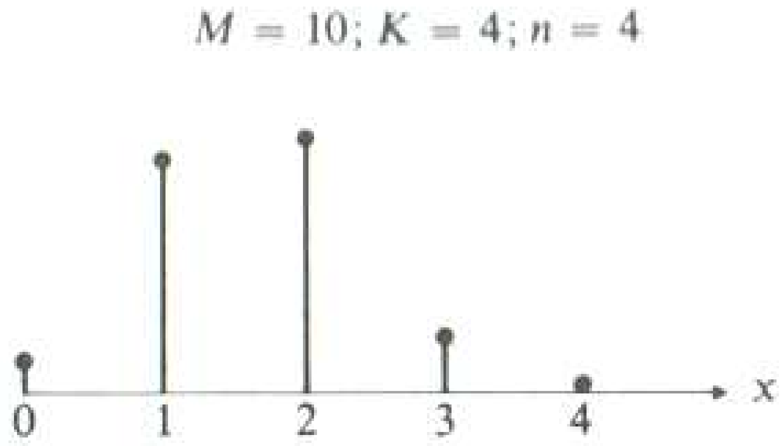
$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

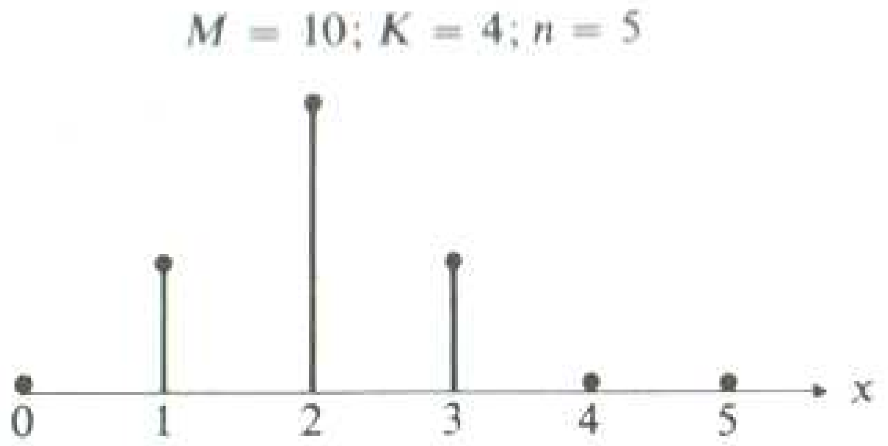
X	$P(X)$
0	${}_nC_0\, p^0 q^n = q^n$
1	${}_nC_1\, p^1 q^{n-1}$
2	${}_nC_2\, p^2 q^{n-2}$
…	…
k	${}_nC_x\, p^k q^{n-k}$
…	…
n	${}_nC_n\, p^n q^0 = p^n$
Total	*1*

The probability that a binomial random variable takes any range of values is the sum of each probability for getting exactly that many successes in *n* observations.

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

2.4 Hypergeometric distribution





Example

Let X denote the number of defectives in a sample of size *n* when sampling is done without replacement from an urn containing *M* balls, *K* of which are defective. Then X has a hypergeometric distribution.

Definition Hypergeometric distribution A random variable X is defined to have a *hypergeometric distribution* if the discrete density function of X is given by

$$f_X(x; M, K, n) = \begin{cases} \dfrac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} & \text{for } x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} I_{\{0, 1, \ldots, n\}}(x)$$

where M is a positive integer, K is a nonnegative integer that is at most M, and n is a positive integer that is at most M. Any distribution function defined by the density function given in Eq. above is called a *hypergeometric distribution.* ////

Theorem If X is a hypergeometric distribution, then

$$\mathscr{E}[X] = n \cdot \frac{K}{M} \quad \text{and} \quad \operatorname{var}[X] = n \cdot \frac{K}{M} \cdot \frac{M-K}{M} \cdot \frac{M-n}{M-1}$$

Proof

$$\mathscr{E}[X] = \sum_{x=0}^{n} x \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} = n \cdot \frac{K}{M} \sum_{x=1}^{n} \frac{\binom{K-1}{x-1}\binom{M-K}{n-x}}{\binom{M-1}{n-1}}$$

$$= n \cdot \frac{K}{M} \sum_{y=0}^{n-1} \frac{\binom{K-1}{y}\binom{M-1-K+1}{n-1-y}}{\binom{M-1}{n-1}}$$

$$= n \cdot \frac{K}{M},$$

using $\sum_{i=0}^{m} \binom{a}{i}\binom{b}{m-i} = \binom{a+b}{m}$

$$\mathscr{E}[X(X-1)]$$

$$= \sum_{x=0}^{n} x(x-1) \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}$$

$$= n(n-1)\frac{K(K-1)}{M(M-1)} \sum_{x=2}^{n} \frac{\binom{K-2}{x-2}\binom{M-K}{n-x}}{\binom{M-2}{n-2}}$$

$$= n(n-1)\frac{K(K-1)}{M(M-1)} \sum_{y=0}^{n-2} \frac{\binom{K-2}{y}\binom{M-2-K+2}{n-2-y}}{\binom{M-2}{n-2}} = n(n-1)\frac{K(K-1)}{M(M-1)}.$$

Hence

$$\begin{aligned}
\text{var}[X] &= \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = \mathscr{E}[X(X-1)] + \mathscr{E}[X] - (\mathscr{E}[X])^2 \\
&= n(n-1)\frac{K(K-1)}{M(M-1)} + n\frac{K}{M} - n^2\frac{K^2}{M^2} \\
&= n\frac{K}{M}\left[(n-1)\frac{K-1}{M-1} + 1 - \frac{nK}{M}\right] \\
&= \frac{nK}{M}\left[\frac{(M-K)(M-n)}{M(M-1)}\right].
\end{aligned}$$

////

Remark

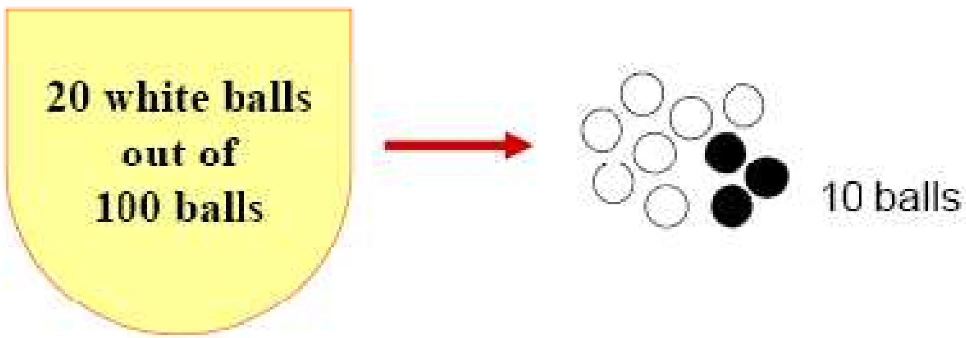
- If we set *K/M=p*, then the mean of the hypergeometric distribution coincides with the mean of the binomial distribution, and the variance of the hypergeometric distribution is *(M-n)/(M-1)* times the variance of the binomial distribution

Example

- Gene Ontology Analysis:
  http://www.livestockgenomics.csiro.au/courses/UAB_Course/S14_GeneOntology.pdf

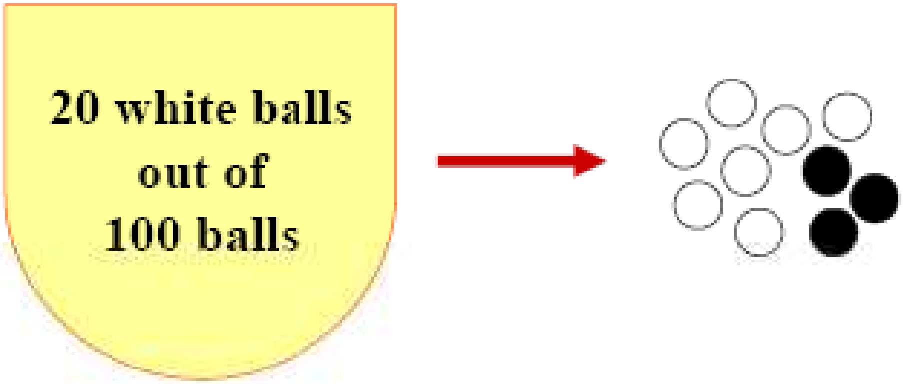
In a given list of genes of interest (eg. DE), is there a Gene Ontology term that is more represented than what it would be expected by chance only?

The hypergeometric distribution arises from sampling from a fixed population.

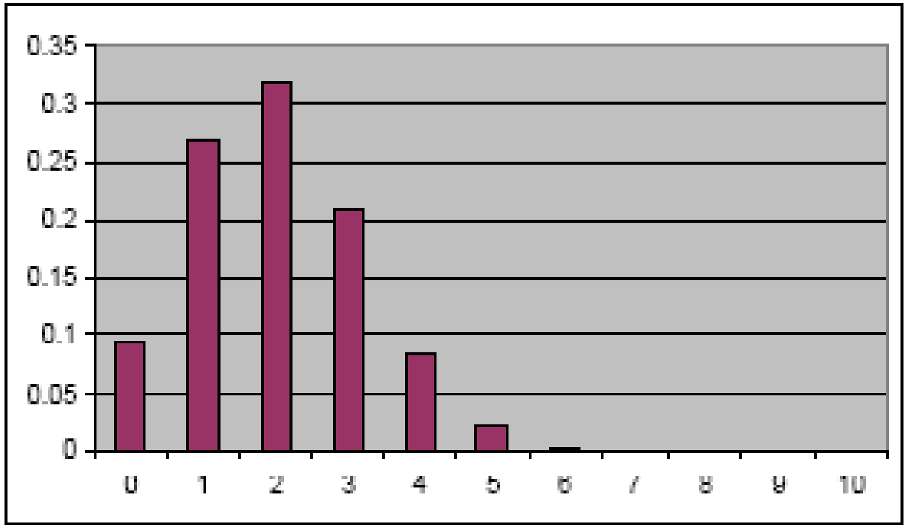


- We want to calculate the probability for drawing 7 or more white balls out of 10 balls given the distribution of balls in the urn

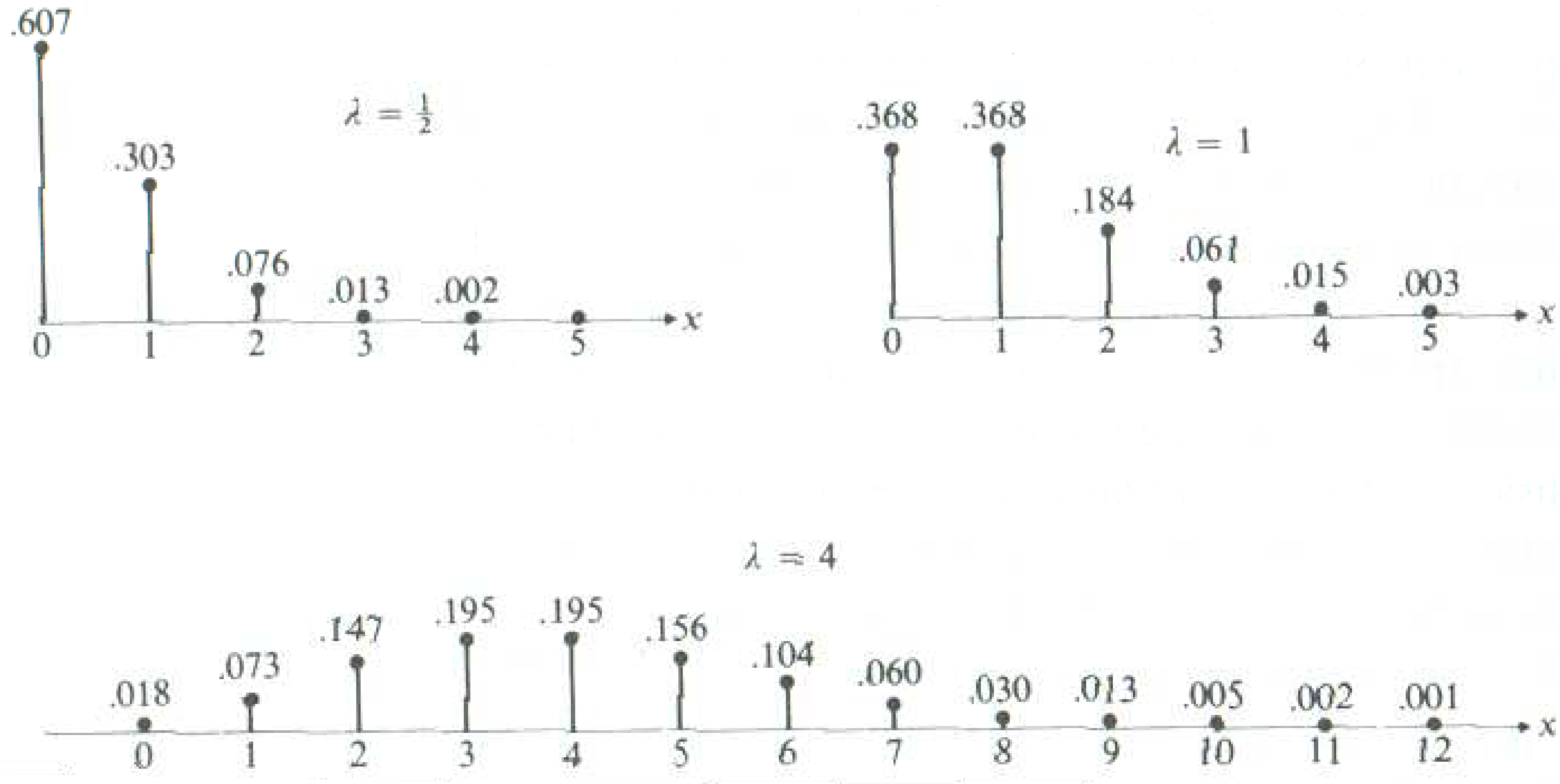
Hypergeometric test … (see later) … to determine whether a GO term is overrepresented or not:



$$P(z,n,t,x)=\frac{\binom{t}{z}\binom{n-t}{x-z}}{\binom{n}{x}}$$



2.5 Poisson distribution



Definition **Poisson distribution** A random variable X is defined to have a *Poisson distribution* if the density of X is given by

$$f_X(x) = f_X(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} = \frac{e^{-\lambda}\lambda^x}{x!} I_{\{0, 1, \ldots\}}(x),$$

where the parameter λ satisfies $\lambda > 0$. The density given in Eq. (9) is called a *Poisson density*. ////

Theorem Let X be a Poisson distributed random variable; then

$$\mathscr{E}[X] = \lambda, \qquad \text{var}\,[X] = \lambda, \qquad \text{and} \quad m_X(t) = e^{\lambda(e^t - 1)}.$$

Proof

$$m_X(t) = \mathscr{E}[e^{tX}] = \sum_{x=0}^{\infty} \frac{e^{tx}e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t};$$

hence,

$$m'_X(t) = \lambda e^{-\lambda} e^t e^{\lambda e^t}$$

and

$$m''_X(t) = \lambda e^{-\lambda} e^t e^{\lambda e^t}[\lambda e^t + 1].$$

So,

$$\mathscr{E}[X] = m'_X(0) = \lambda$$

and

$$\text{var}\,[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = m''_X(0) - \lambda^2 = \lambda[\lambda + 1] - \lambda^2 = \lambda. \qquad ////$$

Common statistics

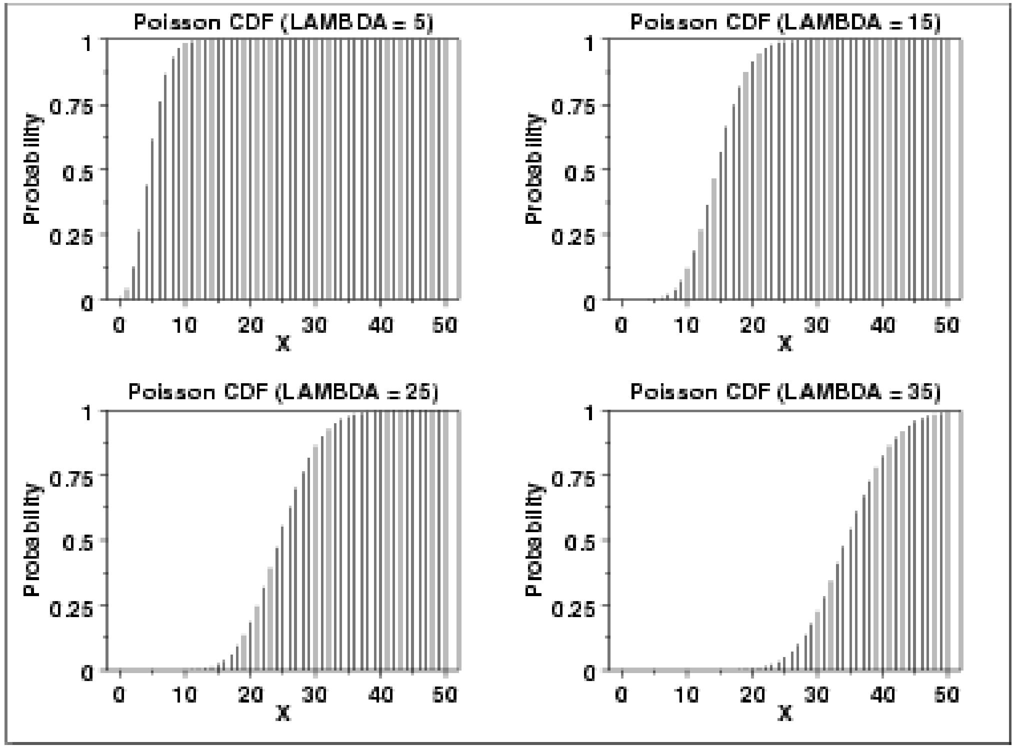
Mean	λ
Mode	For non-integer λ, it is the largest integer less than λ. For integer λ, x = λ and x = λ - 1 are both the mode.
Range	0 to positive infinity
Standard Deviation	$\sqrt{\lambda}$
Coefficient of Variation	$\frac{1}{\sqrt{\lambda}}$
Skewness	$\frac{1}{\sqrt{\lambda}}$
Kurtosis	$3+\frac{1}{\lambda}$

Cumulative distribution function

- The formula for the Poisson cumulative probability function is

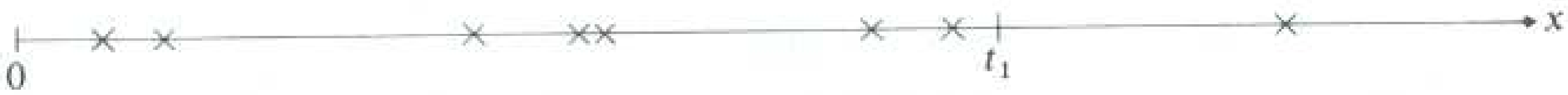
$$F(x, \lambda) = \sum_{i=0}^{x} \frac{e^{-\lambda} \lambda^i}{i!}$$

- The following is the plot of the Poisson cumulative distribution



Example

- The Poisson distribution is used to model the number of events occurring within a given time interval.



- An event or happening may be a fatal traffic accident, a particle emission, a meteorite collision, a flaw in length of a wire, etc, and is denoted by an x in the graph above.
- Now assume that there exists a positive quantity ν, which satisfies the following properties (i) to (iii):

(i) The probability that exactly one happening will occur in a small time interval of length h is approximately equal to νh, or P[one happening in interval of length h] $= \nu h + o(h)$.

(ii) The probability of more than one happening in a small time interval of length h is negligible when compared to the probability of just one happening in the same time interval, or P[two or more happenings in interval of length h] $= o(h)$.

(iii) The numbers of happenings in nonoverlapping time intervals are independent.

- o(h) = "some function of smaller order than *h*":

$$\lim_{h \to 0} \frac{o(h)}{h} = 0$$

- ν can be interpreted as the "mean rate at which events occur per unit of time" and therefore usually referred to as the **mean rate of occurrence**

Theorem If the above three assumptions are satisfied, the number of occurrences of a *happening* in a period of time of length t has a Poisson distribution with parameter $\lambda = \nu t$. Or if the random variable $Z(t)$ denotes the number of occurrences of the happening in a time interval of length t, then $P[Z(t) = z] = e^{-\nu t}(\nu t)^z/z!$ for $z = 0, 1, 2, \ldots$.

Proof *(important)*

For convenience, let t be a point in time after time 0; so the time interval $(0, t]$ has length t, and the time interval $(t, t + h]$ has length h. Let $P_n(s) = P[Z(s) = n] = P$[exactly n happenings in an interval of length s]; then

$$\begin{aligned} P_0(t + h) &= P[\text{no happenings in interval } (0, t + h]] \\ &= P[\text{no happenings in } (0, t] \textit{ and } \text{no happenings in } (t, t + h]] \\ &= P[\text{no happenings in } (0, t]]P[\text{no happenings in } (t, t + h]] \\ &= P_0(t)P_0(h), \end{aligned}$$

using (iii), the independence assumption.

Now P[no happenings in $(t, t+h]] = 1 - P$[one or more happenings in $(t, t+h]] = 1 - P$[one happening in $(t, t+h]] - P$[more than one happening in $(t, t+h]] = 1 - vh - o(h) - o(h)$; so $P_0(t+h) = P_0(t)[1 - vh - o(h) - o(h)]$, or

$$\frac{P_0(t+h) - P_0(t)}{h} = -vP_0(t) - P_0(t)\frac{o(h) + o(h)}{h},$$

and on passing to the limit one obtains the differential equation $P_0'(t) = -vP_0(t)$, whose solution is $P_0(t) = e^{-vt}$, using the condition $P_0(0) = 1$. Similarly, $P_1(t+h) = P_1(t)P_0(h) + P_0(t)P_1(h)$, or $P_1(t+h) = P_1(t)[1 - vh - o(h)] + P_0(t)[vh + o(h)]$, which gives the differential equation $P_1'(t) = -vP_1(t) + vP_0(t)$, the solution of which is given by $P_1(t) = vte^{-vt}$, using the initial condition $P_1(0) = 0$. Continuing in a similar fashion one obtains $P_n'(t) = -vP_n(t) + vP_{n-1}(t)$, for $n = 2, 3, \ldots$.

It is seen that this system of differential equations is satisfied by $P_n(t) = (vt)^n e^{-vt}/n!$.

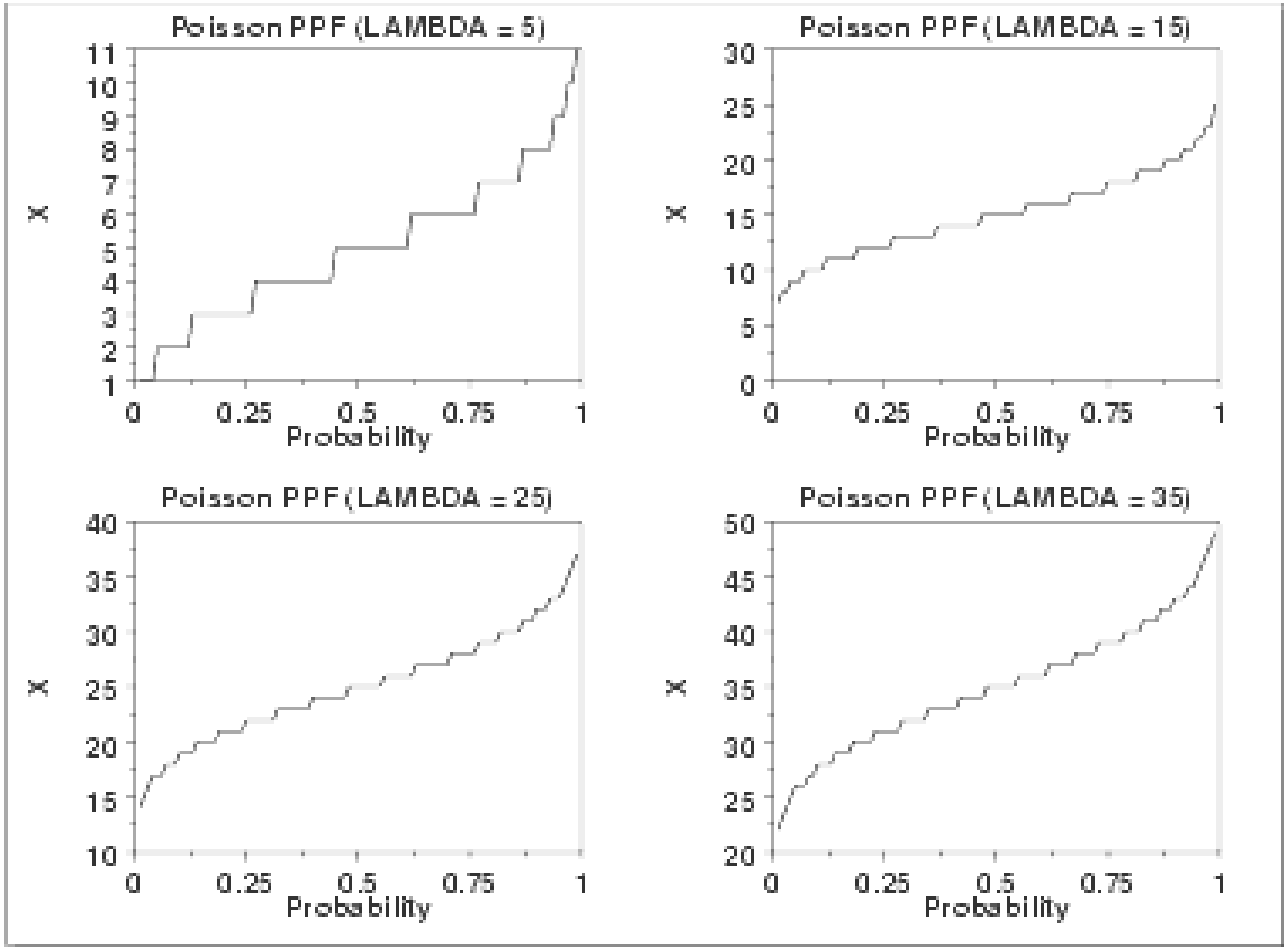
EXAMPLE Suppose that the average number of telephone calls arriving at the switchboard of a small corporation is 30 calls per hour. (i) What is the probability that no calls will arrive in a 3-minute period? (ii) What is the probability that more than five calls will arrive in a 5-minute interval? Assume that the number of calls arriving during any time period has a Poisson distribution. Assume that time is measured in minutes; then 30 calls per hour is equivalent to .5 calls per minute, so the *mean rate of occurrence* is .5 per minute. P[no calls in 3-minute period] $= e^{-\nu t} = e^{-(.5)(3)} = e^{-1.5} \approx .223$.

$$P[\text{more than five calls in 5-minute interval}] = \sum_{k=6}^{\infty} \frac{e^{-\nu t}(\nu t)^k}{k!}$$

$$= \sum_{k=6}^{\infty} \frac{e^{-(.5)(5)}(2.5)^k}{k!} \approx .042. \qquad ////$$

Remark

- The Poisson percent point function does not exist in simple closed form. It is computed numerically.
- Because it is a discrete distribution, it is only defined for integer values of *x*, the percent point function is not smooth in the way the percent point function typically is for a continuous distribution


Poisson PPF (LAMBDA = 5)
Poisson PPF (LAMBDA = 15)
Poisson PPF (LAMBDA = 25)
Poisson PPF (LAMBDA = 35)
X
Probability
0
0.25
0.5
0.75
1