5. SOME IMPORTANT PROBABILITY DISTRIBUTIONS

5.1. Discrete Probability Distributions

Study time: 50 minutes

Aim - you will be able to

- characterize Bernoulli trials and types of discrete distributions
- characterize Poisson process and Poisson distribution
- describe contexture in between discrete distributions

Explication

A lot of discrete random variables exist and now we summarize basic information about the most common discrete variables.
Bernoulli trials:

- a sequence of Bernoulli trials is defined as a sequence of random events which are mutually independent and which have only two possible outcomes (e.g. success-nonsuccess, 1-0)
- probability of event occurrence (a success) p is constant in any trial

$$\mathrm{P}\{\text{Trial 'i'} = \text{"Success"}\} = p$$

Binomial random variable:

The most natural random variable to define on the sample space of Bernoulli trials is the number of successes. Such a random variable is called a binomial random variable. If X is the number of successes in n Bernoulli trials where the probability of success at each trial is p, then we represent the distribution of X by the short-hand notation:

$$X \rightarrow B(n, p)$$

where B indicates that X has a binomial distribution and n and p are the parameters determining which particular distribution from the binomial family applies to X.

The probability distribution of a binomial random variable can be expressed algebraically as:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k};\ 0\le k\le n$$

$$EX=\sum_{k=0}^{n}k\cdot P(X=k)=\sum_{k=0}^{n}k\cdot\frac{n!}{(n-k)!k!}\cdot p^k(1-p)^{n-k}=n\cdot p\sum_{k=0}^{n}\frac{(n-1)!}{(n-k)!\cdot(k-1)!}\cdot p^{k-1}\cdot(1-p)^{n-k}=n\cdot p$$

$DX = EX^2 - (EX)^2$

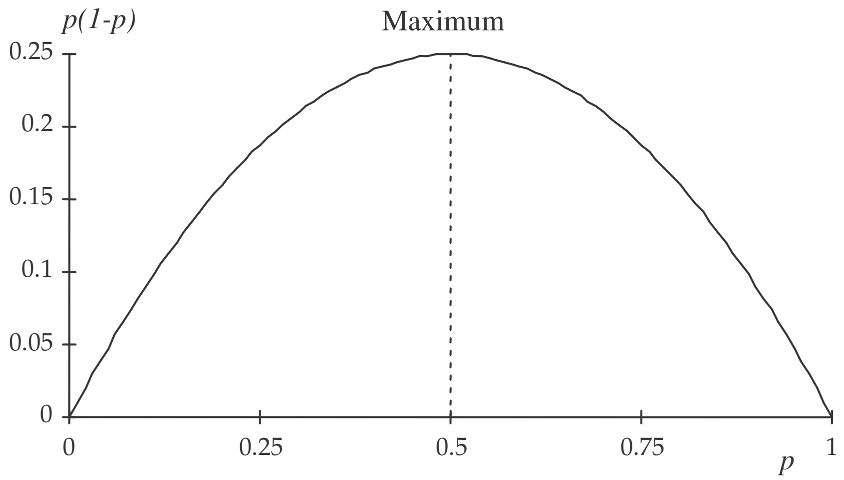
$$EX^2=\sum_{k=1}^{n}k^2\cdot P(X=k)=\sum_{k=1}^{n}k\cdot(k-1)\cdot\frac{n!}{(n-k)!\cdot k!}\cdot p^k\cdot(1-p)^{n-k}+EX=$$

$$=n\cdot(n-1)\cdot p^2\cdot\sum_{k=2}^{n}\frac{(n-2)!}{(n-k)!\cdot(k-2)!}\cdot p^{k-2}\cdot(1-p)^{n-k}+EX=$$

$$=n\cdot(n-1)\cdot p^2+n\cdot p=(np)^2-np^2+np$$

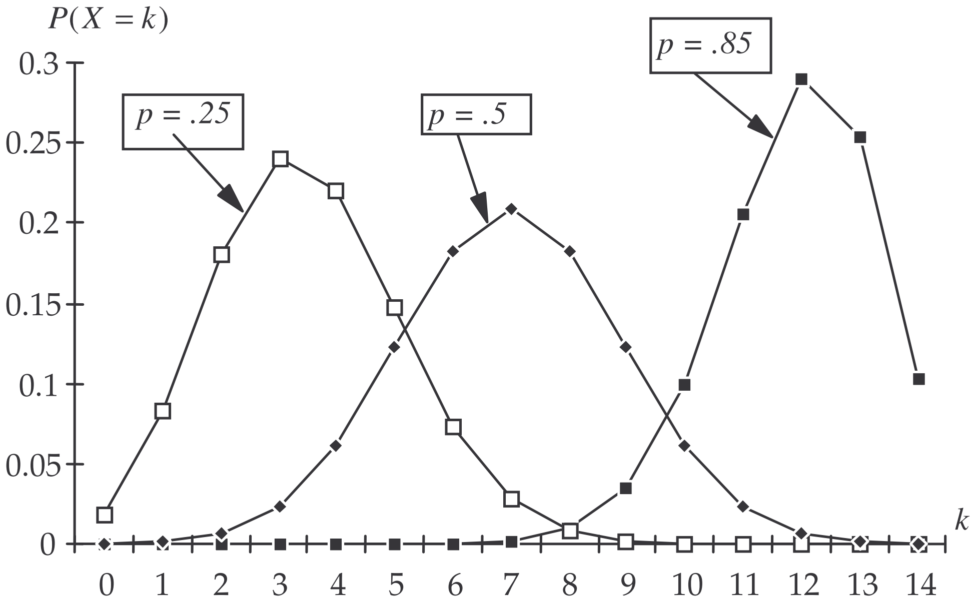
$DX = EX^2 - (EX)^2 = n.p.(1 - p)$

Notice that the variance of the binomial distribution is maximum when $p = 0.5$.



Example:

Some examples of binomial distributions for $n = 14$ trials are illustrated below. Notice that as p, the probability of success at each trial increases, the location of the distribution shifts to higher values of the random variable. Also notice that when p = 0.5, the distribution is symmetric around 7.5.



Geometric distribution:

This distribution has a single parameter, p, and we denote the family of geometric distributions by

$$X \rightarrow G(p)$$

G(p) … the geometric random variable is defined as the number of trials until a success occurs or until the first success

The probability distribution for a geometric random variable is:

$$P(X = k) = p\,(1-p)^{k-1};\ 1 \le k < \infty$$

The expression for the mean of the geometric distribution is

$$EX = \sum_{k=1}^{\infty} k \cdot P(X=k) = \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} = p \cdot \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} = p \cdot \frac{\partial}{\partial(1-p)} \cdot \left(\sum_{k=1}^{\infty} (1-p)^k \right) = \frac{1}{p}$$

Note: By first evaluating the series and then taking its derivative the result is obtained.

The mean number of Bernoulli trials until the first success is the inverse of the success probability at each trial, again an entirely intuitive result. That is, if 10% of the trials are successful, on average it will take ten trials to obtain a success.

To find the variance, we first evaluate the expected value of X^2 and then modify the expression using the same technique as for the binomial case. We note that the expression now has the form of the second derivative of the same geometric series we evaluated for the mean. Taking derivatives of this evaluated expression, we obtain

$$E(X^2) = \sum_{k=1}^{\infty} k^2 \, p(1-p)^{k-1}$$

$$= p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} + p\sum_{k=2}^{\infty} k(1-p)^{k-1}$$

$$= p(1-p)\frac{\partial^2 \sum_{k=1}^{\infty}(1-p)^k}{\partial^2(1-p)} + p\frac{\partial \sum_{k=1}^{\infty}(1-p)^k}{\partial(1-p)}$$

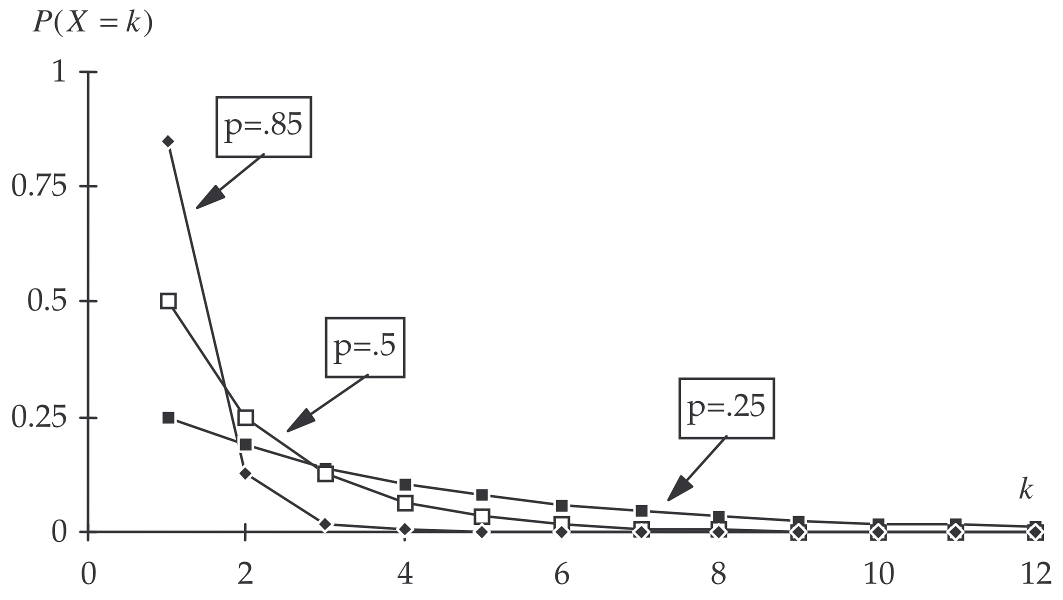
$$= p(1-p)\frac{\partial^2\left(\frac{1-p}{p}\right)}{\partial^2(1-p)} + p\frac{\partial\left(\frac{1-p}{p}\right)}{\partial(1-p)} = \frac{2(1-p)}{p^2} + \frac{1}{p}$$

From the mean and expected value of X^2 we can derive the variance.

$$DX = EX^2 - (EX)^2 = \frac{1-p}{p^2}$$

Example:

Some examples of geometric distributions are illustrated below. Not surprisingly, the probability of long sequences without success decreases rapidly as the success probability, p, increases.



Negative binomial random variable

The negative binomial distribution has two parameters, k and p and is denoted by

$$X \rightarrow NB(k, p)$$

The negative binomial is the number of Bernoulli trials until the k^{th} success.

The negative binomial distribution is:

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

The mean and variance of the negative binomial distribution can be computed easily by noting that a negative binomial random variable with parameters k and p is just the sum of k independent geometric random variables with parameter p. Independence the geometric random variables follow from the independence assumption of sequences of Bernoulli trials. Thus if

$$W_i \rightarrow G(p); 1 \le i \le k$$

then

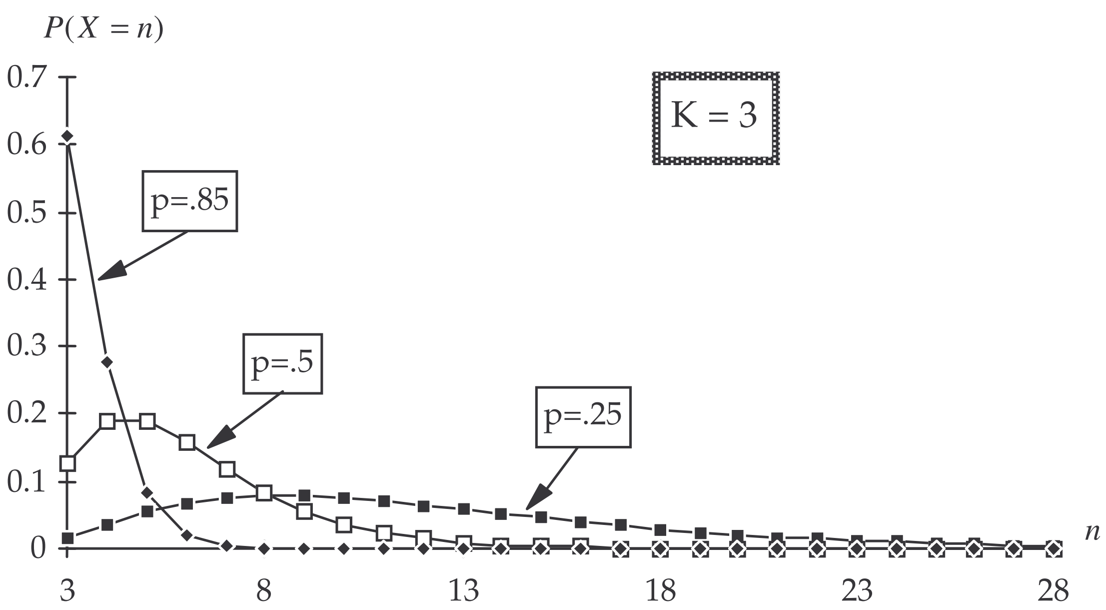
$$X = \sum_{i=1}^{k} W_i$$

$$E(X) = \sum_{i=1}^{k} E(W_i) = \frac{k}{p}$$

$$D(X) = \sum_{i=1}^{k} D(W_i) = \frac{k(1-p)}{p^2}$$

Example:

The following chart illustrates some examples of negative binomial distributions for k=3. Notice that for values of *p* near 0.5, the distribution has a single mode near 5. This mode moves away from the origin and diminishes in magnitude as *p* decreases indicating an increase in variance for small *p*. The negative binomial distribution has a shape similar to the geometric distribution for large values of *p*.



Note:
Comparison of Binomial and Negative Binomial Distributions

It is interesting to compare the distributions of binomial and negative binomial random variables. Notice that except for the combinatorial term at the beginning of each distributional expression, the portion contributed by the probability of single sample space elements is identically $p^k (1 - p)^{n-k}$.

Binomial distribution

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; 0 \le k \le n$$

For the binomial, the number of trials (n) is fixed and the number of successes (k) is variable.

Negative binomial distribution

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

For the negative binomial, the number of successes (k) is fixed and the number of trials (n) is variable.

Poisson process

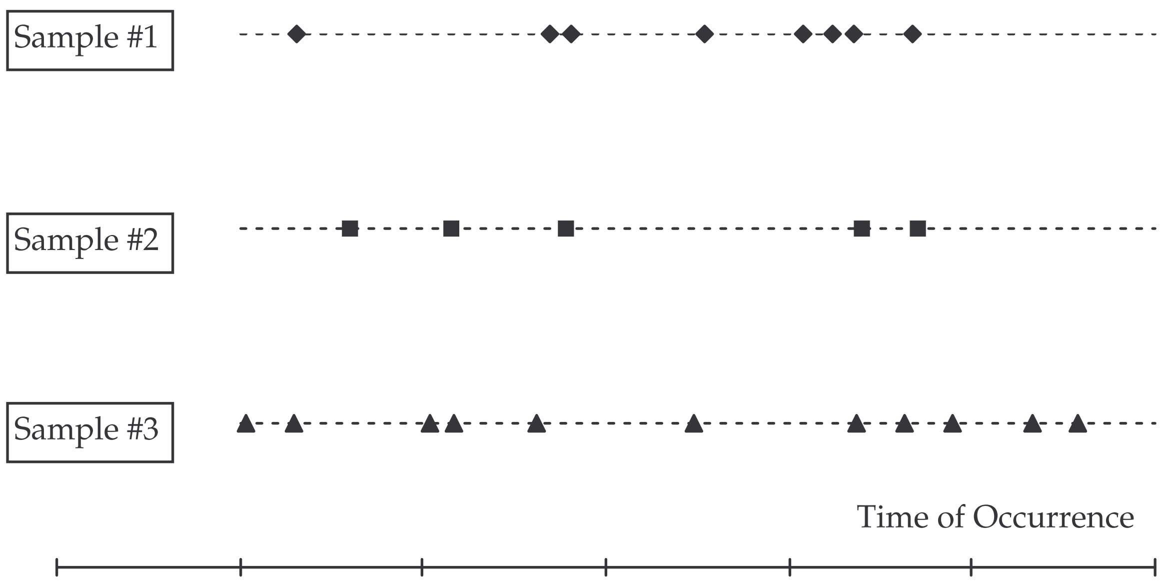
The Poisson process is a second general type of sample space model which is widely applied in practice. The Poisson process may be viewed as the continuous time generalization of a sequence of Bernoulli trials, sometimes called the Bernoulli process. The Poisson process describes the sample space of randomly occurring events in some time interval. The Poisson process assumes that the **rate at which events occur is constant** throughout the interval or region of observation and those events occur **independently** of each other.
λ … lambda → the rate at events occur

However, events occurring over some region can also be modeled by a Poisson process. The appearance of defects in some product, or mold on the surface of a leaf under certain conditions could follow a Poisson process.

Examples:

- Customers arriving at a bank to transact business.
- patients arriving at a clinic for treatment
- telephone inquiries received by a government office, *etc.*

Some examples of elements of the sample space for a Poisson process are illustrated below. This is a complex difficult to characterize sample space. The number of elements in the sample space is uncountable infinite and in this sense continuous. Probabilities cannot be assigned to individual elements of this sample space, only to subsets.



The Poisson process describes events which occur randomly over some time interval or spatial region.

Poisson distribution

The Poisson distribution has a single parameter and therefore we denote this random variable by the symbolic notation,

$$X \rightarrow P(\lambda t)$$

Consider a Poisson process that is observed for a time period t. Suppose the rate of occurrence of events is λ during the time period. Then the total rate of occurrence over the entire observation interval is λt. Now divide the interval t into n subintervals of equal length t/n. Occurrence of events in each of these intervals will be mutually independent at constant rate $\lambda t/n$. If n becomes large enough, the interval lengths, t/n, will become small enough that the probability of two events in one interval is effectively zero and the probability of one event is proportional to $\lambda t/n$. Then the distribution of the number of events in the total interval t can be approximated by the distribution of a binomial random variable with parameters n and $\lambda t/n$. Thus,

$$P(X = k) = \binom{n}{k}\left(\frac{\lambda t}{n}\right)^{k}\left(1-\frac{\lambda t}{n}\right)^{n-k}$$

Taking the limit as n goes to infinity, this expression becomes

$$\lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda t}{n}\right)^{k}\left(1-\frac{\lambda t}{n}\right)^{n-k} = \frac{(\lambda t)^k}{k!}\lim_{n\to\infty}\frac{n!}{(n-k)!\,n^k}\left(1-\frac{\lambda t}{n}\right)^{n-k} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad ,$$

We can express the distribution of a Poisson random variable as:

$$P(X=k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!};\ 0\le k<\infty$$

$$\sum_{k=0}^{\infty}P(X=k)=\sum_{k=0}^{\infty}\frac{(\lambda t)^k e^{-\lambda t}}{k!}=e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}=1$$

Calculation of mean:

$$E(X)=\sum_{k=0}^{\infty}kP(X=k)=\sum_{k=0}^{\infty}k\frac{(\lambda t)^k e^{-\lambda t}}{k!}=\lambda t\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}e^{-\lambda t}}{(k-1)!}=\lambda t$$

The expected value of X^2 is found using the Taylor series expansion of the exponential function as well as the same algebraic manipulation as was invoked for Bernoulli random variables.

$$E(X^2)=\sum_{k=0}^{\infty}k^2P(X=k)=\sum_{k=2}^{\infty}k(k-1)\frac{(\lambda t)^k e^{-\lambda t}}{k!}+\mathrm{E}(\mathrm{X})=$$

$$=(\lambda t)^2\sum_{k=2}^{\infty}\frac{(\lambda t)^{k-2}e^{-\lambda t}}{(k-2)!}+\lambda t=(\lambda t)^2+\lambda t$$

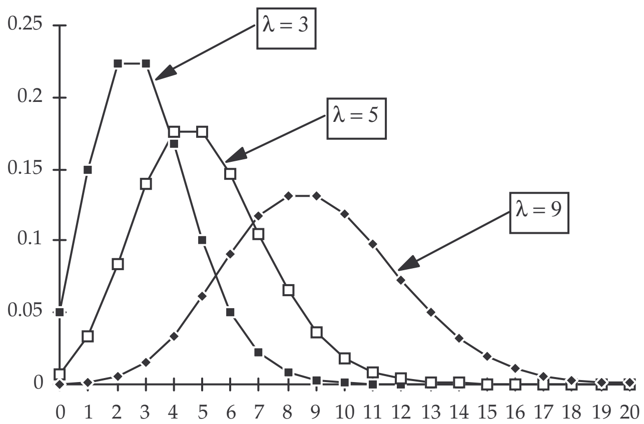
From this result, the variance follows directly.

$$D(X)=E[(X-E(X)]^2)=\mathrm{E}(\mathrm{X}^2)-(\mathrm{E}(\mathrm{X}))^2=\lambda t$$

We notice that the Poisson distribution has the remarkable property that the variance is equal to the mean and by implication that the variance of the Poisson random variable will increase as the rate λ increases.

Example:

Some examples of Poisson distributions are illustrated below. Notice that at the value λ = 9, the distribution becomes almost symmetric.



5.2. Continuous Probability Distributions

Study time: 50 minutes

Aim - you should be able to

- characterize types of continuous distributions : exponential, Gamma and Weibull
- describe contexture in between continuous distributions

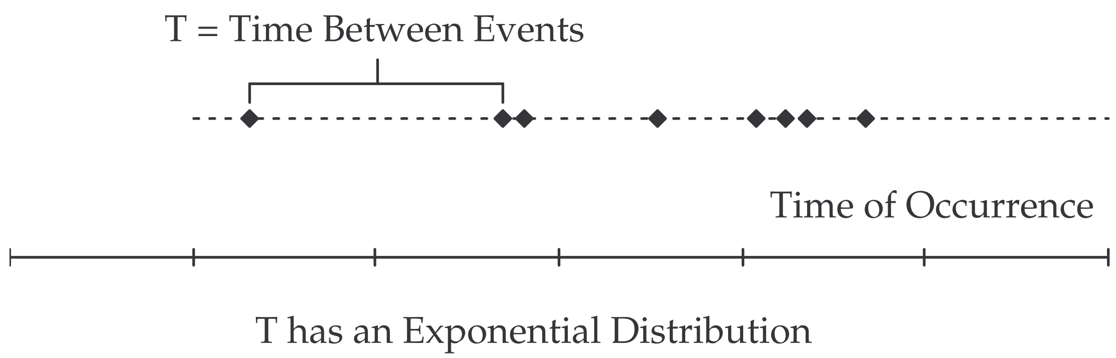
Explication

- **How does a basic description for exponential, gamma and Weibull distributions look like?**

Exponential Distribution

The exponential random variable is a second very natural variable which can be defined on the sample space generated by a Poisson process. If a continuous time process satisfies the assumptions of a Poisson process, the time between events, or equivalently because of the assumption of independence, the time until the next event will have an exponential distribution. The exponential random variable is analogous to the geometric random variable defined for a Bernoulli process.

The range of possible values for the exponential random variable is the set of non-negative numbers.



Strictly speaking, the sample space for an exponential random variable consists of intervals of varying length terminated by a single event, just as the sample space of the geometric random variable consists of a sequence of failures terminated by a success.

The probability density function and distribution function of an exponential distribution have the following simple form.

$$f(t) = \lambda e^{-\lambda t}; t \geq 0$$

$$F(t)=P(T < t) = P(N_t \geq 1) = 1 - P(N_t < 1)$$
$$= 1 - e^{-\lambda t}$$

where λ is the rate at which events occur. The family of exponential random variables is identified by the single parameter, λ, the same parameter which defines the Poisson distribution.

$$T \rightarrow E(\lambda)$$

The mean of the exponential distribution is the reciprocal of the rate parameter. The result can be obtained through integrating the expected value integral by parts.

$$E(T) = \int_{t=0}^{\infty} \lambda t \, e^{-\lambda t} \, dt = \frac{1}{\lambda}$$

The variance of the exponential distribution is obtained from evaluating the following integral again through integration by parts.

$$DT = ET^2 - (ET)^2 = \ldots = \frac{1}{\lambda^2}$$

The variance equals the square of the mean and therefore the mean equals the standard deviation for an exponential distribution.

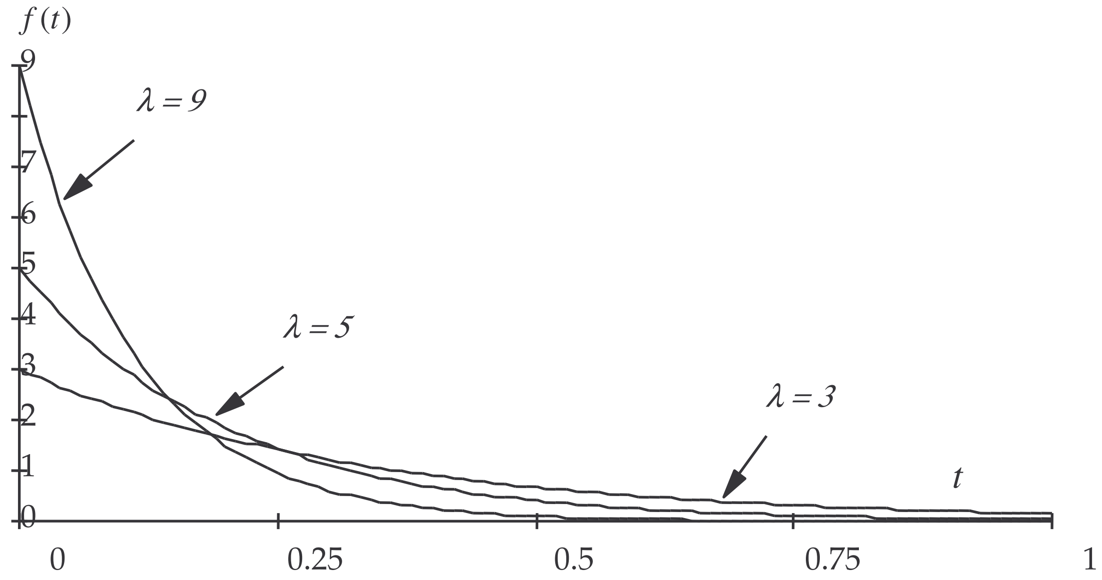
The hazard function is given by:

$$h(t) = \frac{f(t)}{1 - F(t)} \quad \text{if } F(t) < 1$$

$h(t) = \lambda = const. \Rightarrow$ the "no memory" property of the exponential distribution

Example:

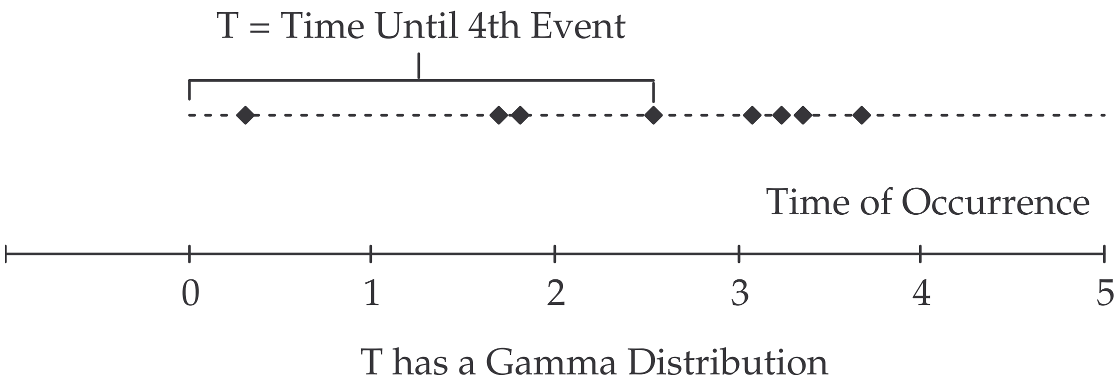
The following graph illustrates some examples of the probability density functions of exponential random variables. Notice that the shape of the exponential density is similar to the shape of the geometric probability distribution. The exponential distribution of time to next event is the continuous time equivalent of the geometric distribution which is the number of trials to next event where event may be considered a "successful" trial.



Gamma distribution

$T \rightarrow Ga(k, \lambda)$

The sample space generated by the Poisson process gives rise to a second random variable closely associated with the exponential random variable. The total time until some specified number of events, say *k*, occur is called a Gamma random variable and arises as a sum of *k* identical independent exponential random variables. If the exponential distribution is the continuous time equivalent of the geometric, then it follows that the Gamma distribution is the continuous time equivalent of the negative binomial.



The sample of a gamma random variable arising as the sum of 4 independent exponential random variables, that is as the time until the fourth event in a Poisson process will consist of intervals of varying length, all having three events and terminated by a fourth event.

The gamma distribution function for any integer value of *k* can be derived by the following argument. Since the gamma arises as the sum of *k* independent, identically distributed exponential random variables, the distribution function of the gamma is the probability that the sum of *k* exponentials is less than or equal to some value *t*. This implies that there have been at least *k* occurrences of a Poisson process within time *t*, the probability of which is given by the cumulative distribution of a Poisson random variable with rate parameter λt, where λ is the rate of the underlying Poisson process.

$$T_k = X_1 + X_2 + X_3 + \ldots + X_k$$

$$X_i \to E(\lambda)$$

$$F(t) = P(T_k < t) = P\left(\sum_{i=1}^{k} X_i < t\right) = P(N_t \geq k) = 1 - P(N_t < k) = 1 - \sum_{j=0}^{k-1} e^{-\lambda t} \cdot \frac{(\lambda t)^j}{j!} \; = 1 - e^{-\lambda t}\left[\sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!}\right]$$

and the probability density function is

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t}\left[\sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!}\right] - e^{-\lambda t}\left[\sum_{j=1}^{k-1} \frac{\lambda(\lambda t)^{j-1}}{(j-1)!}\right] \\ &= \lambda^k e^{-\lambda t}\left[\frac{t^{k-1}}{(k-1)!}\right]; \quad t > 0 \end{aligned}$$

Since the gamma random variable is the sum of k identical independent exponential random variables, the mean and variance will be k times the mean and variance of an exponential random variable. This same argument was used to derive the mean and variance of the negative binomial from the moments of the geometric distribution.

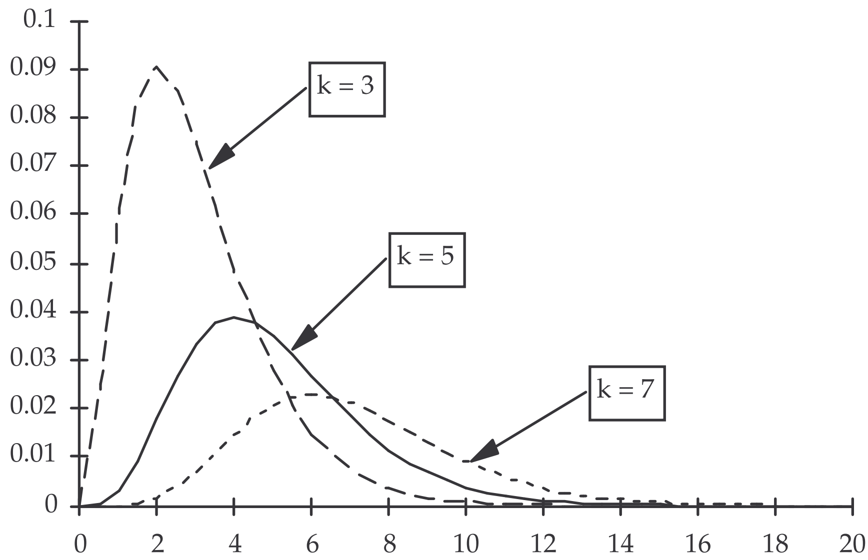
$$ET_k = EX_1 + EX_2 + \ldots + EX_k = \frac{1}{\lambda} + \ldots + \frac{1}{\lambda} = \frac{k}{\lambda}$$

$$DT_k = \ldots. = \frac{k}{\lambda^2}$$

The form of the gamma distribution presented here where the parameter k is restricted to be a positive integer is actually a special case of the more general family of gamma distributions where k is a shape parameter which need only be a positive real number. The special case we have discussed is sometimes called the **Erlang distribution**.

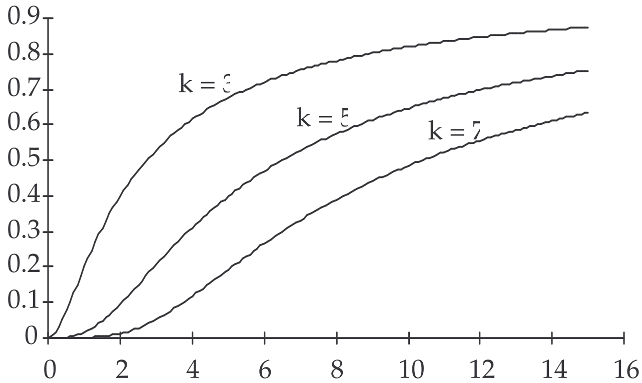
Example:

Examples of gamma probability density functions for $\lambda = 1$ are illustrated in the following chart. Notice that the gamma density has a single mode which moves away from the origin as k increases. Also the dispersion increases and the distribution becomes more symmetric and k increases.



The hazard function is given by:

$$h(x) = \frac{\lambda}{(k-1)!\sum_{j=0}^{k-1} \frac{1}{(k-1-j)!(\lambda x)^j}}$$



The hazard function of Gamma distribution, λ=1

h(x) is a sharply increasing function for k > 1 ⇒ this distribution is suitable for modeling of ageing and wear processes

Weibull distribution

The distribution function is:

$F(x) = 1 - e^{-\left(\frac{x}{\Theta}\right)^{\beta}}$, $\Theta > 0, \beta > 0, x > 0$ β ... a shape parameter, Θ ... a scale parameter.

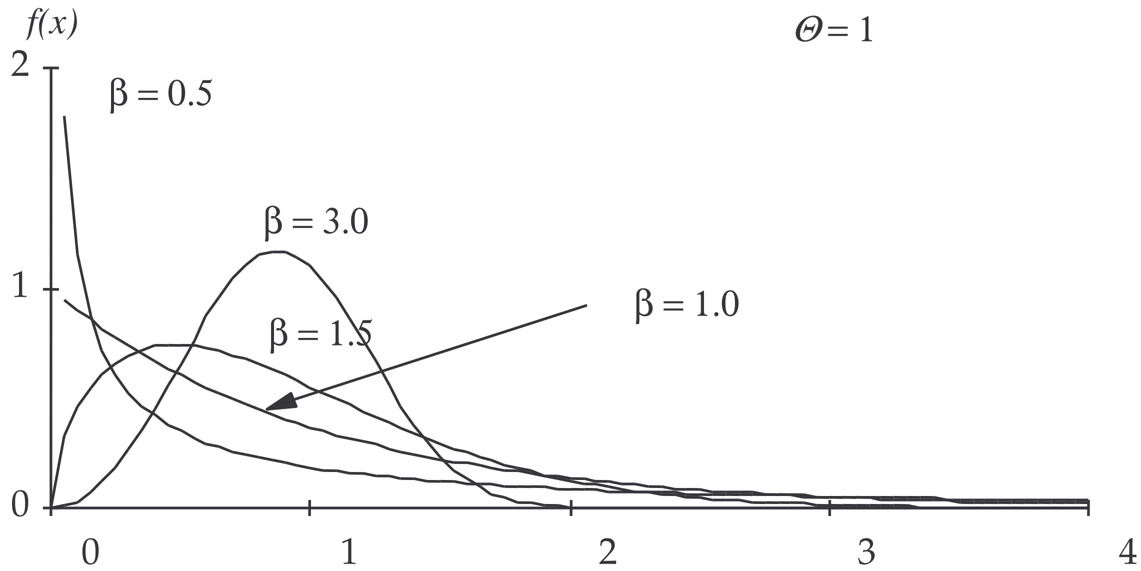
The probability density function for the Weibull is:

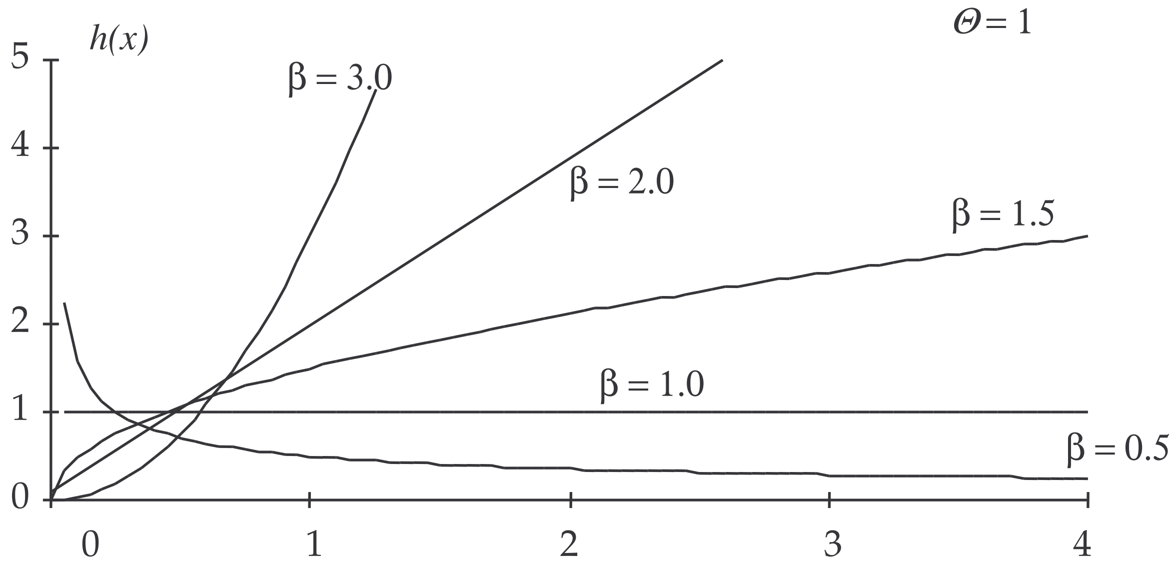
$$f(x) = \frac{\beta}{\Theta}\left(\frac{x}{\Theta}\right)^{\beta-1} e^{-\left(\frac{x}{\Theta}\right)^{\beta}}$$

And the hazard function for the Weibull is:

$$h(x) = \frac{\beta}{\Theta}.\left(\frac{x}{\Theta}\right)^{\beta-1}$$

Some examples of the Weibull density and the Weibull hazard function are illustrated below.





The Weibull distribution is very flexible and we use it in Reliability theory for modeling of the random variable "time to failure".

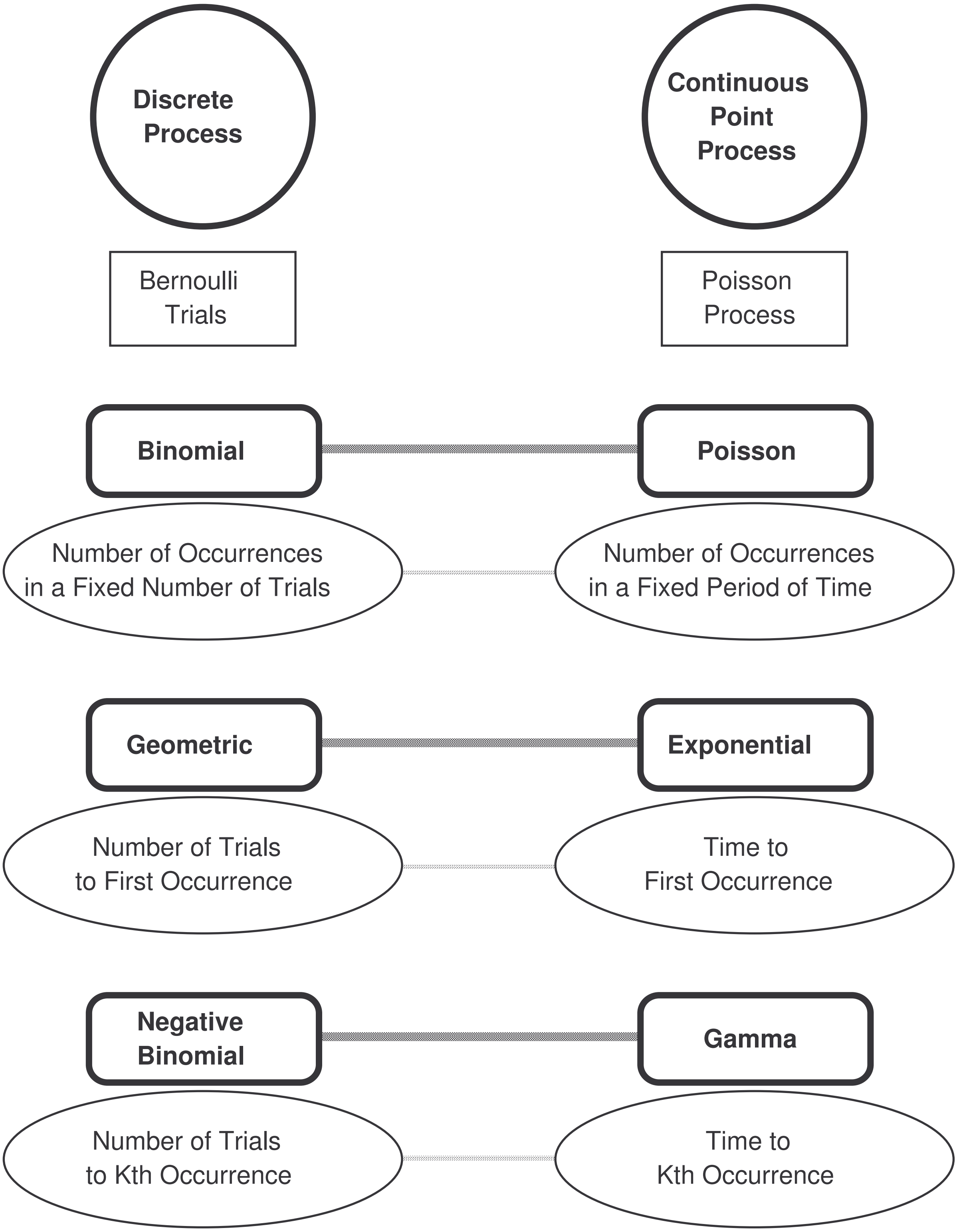
Σ Summary of notions

A sequence of **Bernoulli trials** is defined as a sequence of random events which are mutually independent and which have only two possible outcomes (e.g. success-nonsuccess, 1-0) and the probability of event occur (a success) p is constant in any trial.
On the basis of these trials expectations we can define the following random variables: **binomial**, **geometric** and **negative binomial**.
A number of events occurrences on any deterministic interval from 0 to t can be describe (at certain expectations) by a **Poisson distribution**.
If a continuous time process satisfies the assumptions of a Poisson process, the time between events, or equivalently because of the assumption of independence, the time until the next event will have an **Exponential distribution**.
A Gamma distribution describes a time to k-th event occurrence.
A Weibull distribution is generalization of the exponential distribution and it is very flexible.

? Questions

1. Which discrete and continuous distributions you know?
2. Characterize the Bernoulli trials and individual types of the discrete distributions. Determine a mean of the Binomial random variable.
3. What is Gamma distribution used for? How is it related to Exponential distribution?
4. For what β of Weibull distribution is the hazard function linear increasing?

Diagram: a contexture among distributions



Problems

Example 1*:* Suppose that a coin with probability of heads $p = 0.4$ is tossed 5 times. Let X denote the number of heads.

a) Compute the density function of X explicitly.
b) Identify the mode.
c) Find $P(X > 3)$.

{*Answer:* Let $f(k) = P(X = k) = \binom{5}{k}(0.4)^k\,(0.6)^{5-k}$ for $k = 0, 1, 2, 3, 4, 5$.

a) $f(0) = 0.0778, f(1) = 0.2592, f(2) = 0.3456, f(3) = 0.2304, f(4) = 0.0768, f(5) = 0.0102$.
b) mode: $k = 2$
c) $P(X > 3) = 0.9870$. }

Example 2*:* Suppose that the number of misprints N on a web page has the Poisson distribution with parameter 2.5.

a) Find the mode.
b) Find $P(N > 4)$.

{*Answer:* a) mode: $n = 2$, b)$P(N > 4) = 0.1088$}

Example 3: Message arrive at a computer at an average rate of 15 messages/second. The number of messages that arrive in 1 second is known to be a Poisson random variable.

a) Find the probability that no messages arrive in 1 second.
b) Find the probability that more than 10 messages arrive in a 1-second period.

{*Answer:* a) $3.06(10^{-7})$, b) 0.8815}

Example 4: If there are 500 customers per eight-hour day in a check-out lane, what is the probability that there will be exactly 3 in line during any five-minute period?

{*Answer:* Poisson - 0.1288 }