STK2130 – Week 10

A. B. Huseby

Department of Mathematics
University of Oslo, Norway

Chapter 5

The Exponential Distribution and the Poisson Process

Chapter 5.2 The Exponential Distribution

A continuous random variable X is said to have an exponential distribution with parameter $\lambda > 0$, denoted as $X \sim exp(\lambda)$, if its probability density function is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

If $X \sim exp(\lambda)$, then the cumulative distribution function of X is given by:

$$F(x) = P(X \leq x) = \int_0^x f(t)dt = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Moreover, the survival function of X is given by:

$$\bar{F}(x) = P(X > x) = 1 - F(x) = \begin{cases} e^{-\lambda x} & x \geq 0 \\ 1 & x < 0 \end{cases}$$

Memoryless stochastic variables

A random variable X is said to be memoryless if:

$$P(X > s + t | X > t) = P(X - t > s | X > t) = P(X > s), \quad \text{for all } s, t \geq 0.$$

Thus, X is memoryless if $(X - t)|(X > t)$ has the same distribution as X.

Note that if X is the lifetime of some unit, $(X - t)$ is the remaining lifetime given that the unit has survived up to the time t.

If $X \sim exp(\lambda)$, we have:

$$P(X > s + t | X > t) = \frac{P(X > s + t \cap X > t)}{P(X > t)} = \frac{P(X > s + t)}{P(X > t)}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda(t)}} = e^{-\lambda s} = P(X > s)$$

Hence, we conclude that X is memoryless.

Memoryless stochastic variables (cont.)

The memoryless property:

$$P(X > s + t | X > t) = P(X > s), \quad \text{for all } s, t \geq 0.$$

is equivalent to the following:

$$P(X > s + t) = P(X > s)P(X > t), \quad \text{for all } s, t \geq 0.$$

Since $\bar{F}(x) = P(X > x)$, this property can also be written as:

$$\bar{F}(s + t) = \bar{F}(s)\bar{F}(t), \quad \text{for all } s, t \geq 0.$$

We now show that the exponential distribution is essentially the only distribution with this property.

Memoryless stochastic variables (cont.)

Proposition

Let X be a random variable and let $\bar{F}(x) = P(X > x)$ be such that:

$$\bar{F}(x+y) = \bar{F}(x) \cdot \bar{F}(y), \qquad \textit{for all } x, y \geq 0. \tag{1}$$

$$\lambda = -\log(\bar{F}(1)) > 0. \tag{2}$$

Then $X \sim exp(\lambda)$.

PROOF: We first note that by (2), it follows that:

$$0 < \bar{F}(1) = e^{-\lambda} < 1. \tag{3}$$

Secondly we note that since cumulative distribution functions always are right-continuous, it follows that $\bar{F} = 1 - F$ is righ-continuous as well.

Memoryless stochastic variables (cont.)

By repeated use of (1) it follows that for $n, m \in \mathbb{N}^+$, we have:

$$\bar{F}(\tfrac{m}{n}) = \bar{F}(\tfrac{1}{n} + \cdots + \tfrac{1}{n}) = \bar{F}^m(\tfrac{1}{n}), \tag{4}$$

where the sum contains m terms. In particular, by letting $m = n$, we get:

$$\bar{F}(1) = \bar{F}(\tfrac{n}{n}) = \bar{F}^n(\tfrac{1}{n}). \tag{5}$$

Alternatively, (5) can be written as:

$$\bar{F}(\tfrac{1}{n}) = [\bar{F}(1)]^{1/n}. \tag{6}$$

By (3) and that $\bar{F}$ is right-continuous, (6) implies that:

$$\bar{F}(0) = \lim_{n\to\infty} \bar{F}(\tfrac{1}{n}) = \lim_{n\to\infty} [\bar{F}(1)]^{1/n} = 1.$$

Hence, since $\bar{F}$ must be non-increasing, $\bar{F}(x) = 1$ for all $x \leq 0$.

Memoryless stochastic variables (cont.)

We now combine (4) and (6), and get:

$$\bar{F}(\tfrac{m}{n}) = \bar{F}^m(\tfrac{1}{n}) = \bar{F}(1)^{m/n}, \quad \text{for all } m, n \in \mathbb{N}^+.$$

Thus, since $\bar{F}(1) = e^{-\lambda}$, we have proved that:

$$\bar{F}(q) = \bar{F}(1)^q = e^{-\lambda q}, \quad \text{for all } q \in \mathbb{Q}^+.$$

Now, let $x \in \mathbb{R}^+$. Since the set $\mathbb{Q}^+$ is dense in $\mathbb{R}^+$, there exists a decreasing sequence $\{q_r\} \subset \mathbb{Q}^+$ such that:

$$\lim_{r \to \infty} q_r = x$$

Since $\bar{F}$ is right-continuous, this implies that:

$$\bar{F}(x) = \lim_{r \to \infty} \bar{F}(q_r) = \lim_{r \to \infty} e^{-\lambda q_r} = e^{-\lambda x}.$$

Hence, we conclude that $X \sim exp(\lambda)$ ■

The failure rate function

Let X be a continuously distributed positive random variable with density f, and survival function $\bar{F}(x) = P(X > x)$, where $\bar{F}(0) = 1$.

The failure rate function (or hazard rate function) of X is defined as:

$$r(t) = \frac{f(t)}{\bar{F}(t)}, \quad \text{for all } t \text{ such that } \bar{F}(t) > 0.$$

The failure rate function, $r(t)$ can be interpreted as follows:

$$\begin{aligned} P(X \in (t, t+dt) | X > t) &= \frac{P(X \in (t, t+dt) \cap X > t)}{P(X > t)} \\ &= \frac{P(X \in (t, t+dt))}{P(X > t)} \approx \frac{f(t)dt}{\bar{F}(t)} = r(t)dt. \end{aligned}$$

Thus, $r(t)$ is the conditional probability density given that $X > t$.

The failure rate function (cont.)

Given the function $r(t)$ we can recover the survival function $\bar{F}(x)$:

$$\int_0^x r(t)dt = \int_0^x \frac{f(t)}{\bar{F}(t)}dt \qquad \text{Subst.: } u = F(t),\ du = f(t)dt$$

$$= \int_0^{F(x)} \frac{du}{1-u} = -\Big|_0^{F(x)} [\ln(1-u) + C] = -\ln(\bar{F}(x)) - C.$$

By using that $\bar{F}(0) = 1$, it follows that $C = 0$, and hence we get:

$$\bar{F}(x) = \exp\left(-\int_0^x r(t)dt\right) = e^{-R(x)}, \quad x \geq 0,$$

where $R(x) = \int_0^x r(t)dt$ is referred to as the cumulative failure rate function (or cumulative hazard function).

The failure rate function (cont.)

Assume that $X \sim exp(\lambda)$. Then it follows that:

$$r(t) = \frac{f(t)}{\bar{F}(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda,$$

$$R(x) = \int_0^x r(t)dt = \int_0^x \lambda dt = \lambda x.$$

Hence, for the exponential distribution, the failure rate function is constant, while cumulative failure rate function is linear.

The Weibull distribution

Assume that the random variable X has a cumulative failure rate function $R(x) = (\lambda x)^{\alpha}$ where $\alpha, \lambda > 0$, $x \geq 0$.

Then the distribution of X is called a Weibull distribution. The parameter α is referred to as the shape parameter.

The failure rate function of a Weibull distribution is obtained as the derivative of $R(x)$:

$$r(t) = R'(t) = \alpha\lambda(\lambda t)^{\alpha-1}, \quad t > 0.$$

We observe that:

- If $0 < \alpha < 1$, then r is decreasing
- If $\alpha = 1$, then r is constant
- If $\alpha > 1$, then r is increasing

In fact, if $\alpha = 1$, $X \sim exp(\lambda)$.

The Weibull distribution (cont.)

The survival function of a Weibull distribution is given by:

$$\bar{F}(x) = P(X > x) = e^{-R(x)} = e^{-(\lambda x)^{\alpha}}, \quad x \geq 0.$$

The cumulative distribution function of a Weibull distribution is given by:

$$F(x) = 1 - \bar{F}(x) = 1 - e^{-(\lambda x)^{\alpha}}, \quad x \geq 0.$$

The density function of a Weibull distribution is given by:

$$f(x) = r(x)\bar{F}(x) = \alpha\lambda(\lambda t)^{\alpha-1} e^{-(\lambda x)^{\alpha}}, \quad x \geq 0.$$

Example 5.6 Hyperexponential distribution

Let $X_1, \ldots, X_n$ be independent exponential random variables with respective rates $\lambda_1, \ldots, \lambda_n$, where $\lambda_i \neq \lambda_j$ when $i \neq j$.

Let J be a discrete variable independent of $X_1, \ldots, X_n$ and with values in the index set $\{1, \ldots, n\}$. Assume that:

$$P(J = j) = P_j, \quad j = 1, \ldots, n, \quad \sum_{j=1}^{n} P_j = 1.$$

The random variable X_J is said to a hyperexponential random variable, or a discrete mixture of exponentials.

EXAMPLE: A collection of n types of batteries.

– The lifetime of a battery of type j is $exp(\lambda_j)$, $\quad j = 1, \ldots, n$.

– P_j is the proportion of batteries of type j, $\quad j = 1, \ldots, n$.

Example 5.6 Hyperexponential distribution (cont.)

The cumulative distribution function is given by:

$$\begin{aligned} F(x) &= P(X_J \leq x) = 1 - P(X_J > x) \\ &= 1 - \sum_{j=1}^{n} P(X_j > x | J = j) P(J = j) \\ &= 1 - \sum_{j=1}^{n} P_j e^{-\lambda_j x}, \quad x \geq 0. \end{aligned}$$

The probability density function is given by:

$$f(x) = \frac{d}{dx} F(x) = \sum_{j=1}^{n} P_j \lambda_j e^{-\lambda_j x}, \quad x \geq 0.$$

Example 5.6 Hyperexponential distribution (cont.)

The failure rate function is given by:

$$r(t) = \frac{f(t)}{\bar{F}(t)} = \frac{\sum_{j=1}^{n} P_j \lambda_j e^{-\lambda_j t}}{\sum_{j=1}^{n} P_j e^{-\lambda_j t}}, \quad t > 0.$$

NOTE: We also have:

$$P(J = j | X > t) = \frac{P(X > t | J = j) P(J = j)}{P(X > t)} = \frac{P_j e^{-\lambda_j t}}{\sum_{j=1}^{n} P_j e^{-\lambda_j t}}, \quad t > 0.$$

Hence, the failure rate function $r(t)$ can alternatively be expressed as:

$$r(t) = \sum_{j=1}^{n} \lambda_j P(J = j | X > t).$$

Example 5.6 Hyperexponential distribution (cont.)

If $\lambda_1 < \lambda_i$ for all $i > 1$, then:

$$P(J = 1|X > t) = \frac{P_1 e^{-\lambda_1 t}}{P_1 e^{-\lambda_1 t} + \sum_{j=2}^{n} P_j e^{-\lambda_j t}}$$

$$= \frac{P_1}{P_1 + \sum_{j=2}^{n} P_j e^{-(\lambda_j - \lambda_1)t}} \to 1, \qquad \text{as } t \to \infty.$$

Similarly, for all $j > 1$ we have:

$$P(J = j|X > t) \to 0, \qquad \text{as } t \to \infty.$$

From this it follows that:

$$\lim_{t\to\infty} r(t) = \min_{1 \le j \le n} \lambda_j.$$

The Exponential Distribution (cont.)

Proposition (5.2)

Assume that $X_1, \ldots, X_n$ are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, \ldots, n$. Then $\min_i X_i \sim exp(\sum_{i=1}^n \lambda_i)$. Moreover, $\min_i X_i$ and the rank order of $X_1, \ldots, X_n$ are independent.

PROOF: Since the exponential distribution is memoryless, we get that:

$$
\begin{aligned}
P(X_{i_1} < \cdots < X_{i_n} \mid \min_{1 \leq i \leq n} X_i > t) \\
&= P(X_{i_1} < \cdots < X_{i_n} \mid \bigcap_{i=1}^n X_i > t) \\
&= P(X_{i_1} - t < \cdots < X_{i_n} - t \mid \bigcap_{i=1}^n X_i > t) \\
&= P(X_{i_1} < \cdots < X_{i_n})
\end{aligned}
$$

Example 5.8

A post office with two clerks, both are busy but there is no one else waiting in line.

$$R_i = \text{Time until for clerk } i \text{ becomes available}, \quad i = 1, 2$$

$$S = \text{Your service time}$$

$$T = \text{The total time spent in the post office}$$

We assume that R_1, R_2 are independent and $R_i \sim exp(\lambda_i)$, $i = 1, 2$.

$$E[T] = E[T | R_1 < R_2] P(R_1 < R_2) + E[T | R_2 \leq R_1] P(R_2 < R_1)$$

$$= E[R_1 + S | R_1 < R_2] \frac{\lambda_1}{\lambda_1 + \lambda_2} + E[R_2 + S | R_2 < R_1] \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

We now use that:

$$E[R_i | R_i < R_{3-i}] = E[\min(R_1, R_2)] = \frac{1}{\lambda_1 + \lambda_2}, \quad i = 1, 2$$

$$E[S | R_i < R_{3-i}] = \frac{1}{\lambda_i}, \quad i = 1, 2$$

Example 5.8 (cont.)

$$
\begin{aligned}
E[T] &= E[R_1 + S | R_1 < R_2] \frac{\lambda_1}{\lambda_1 + \lambda_2} + E[R_2 + S | R_2 < R_1] \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
&= \left(\frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_1} \right) \frac{\lambda_1}{\lambda_1 + \lambda_2} + \left(\frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_2} \right) \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
&= \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} + 1 \right) \frac{1}{\lambda_1 + \lambda_2} + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} + 1 \right) \frac{1}{\lambda_1 + \lambda_2} \\
&= \left(\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2} + 1 + 1 \right) \frac{1}{\lambda_1 + \lambda_2} \\
&= \frac{3}{\lambda_1 + \lambda_2}.
\end{aligned}
$$

The hypoexponential distribution

We recall that if $X \sim exp(\lambda)$, then the moment generating function of X is given by:

$$M_X(t) = E[e^{tX}] = \int_0^\infty \lambda e^{-(\lambda - t)x} dx = \frac{\lambda}{\lambda - t}.$$

Now, let $X_1, \ldots, X_n$ be independent and $X_i \sim exp(\lambda_i)$, $i = 1, \ldots, n$, and assume that all the λ_i's are distinct. That is $\lambda_i \neq \lambda_j$ for all $i \neq j$.

The moment generating function of $S = X_1 + \cdots + X_n$ is given by:

$$\begin{aligned}
M_S(t) &= E[e^{tS}] = E[e^{tX_1 + \cdots tX_n}] \\
&= E[e^{tX_1}] \cdots E[e^{tX_n}] \quad \text{(since } X_1, \ldots, X_n \text{ are independent)} \\
&= \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n \frac{\lambda_i}{\lambda_i - t}.
\end{aligned}$$

The hypoexponential distribution (cont.)

Assume that $\lambda_1, \ldots, \lambda_n$ be distinct positive numbers. A random variable Z is said to have a hypoexponential distribution with rates $\lambda_1, \ldots, \lambda_n$ if the density of Z is given by:

$$f_Z(z) = \sum_{i=1}^{n} C_{i,n} \cdot \lambda_i e^{-\lambda_i z}, \quad z \geq 0,$$

where:

$$C_{i,n} = \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i}, \quad i = 1, \ldots, n.$$

The moment generating function of Z is then given by:

$$M_Z(t) = E[e^{tZ}] = \int_0^\infty \sum_{i=1}^{n} C_{i,n} \cdot \lambda_i e^{-(\lambda_i - t)z} dz$$

$$= \sum_{i=1}^{n} C_{i,n} \int_0^\infty \lambda_i e^{-(\lambda_i - t)z} dz = \sum_{i=1}^{n} C_{i,n} \cdot \frac{\lambda_i}{\lambda_i - t}$$

The hypoexponential distribution (cont.)

By inserting the expressions for $C_{1,n}, \ldots, C_{n,n}$, we get:

$$M_Z(t) = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i - t} \cdot C_{i,n} \; = \; \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i - t} \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

$$= \prod_{i=1}^{n} \frac{\lambda_i}{\lambda_i - t} \cdot \sum_{i=1}^{n} \prod_{j \neq i} \frac{\lambda_j - t}{\lambda_j - \lambda_i} \; = \; \prod_{i=1}^{n} \frac{\lambda_i}{\lambda_i - t} \cdot \phi_n(t),$$

where:

$$\phi_n(t) = \sum_{i=1}^{n} \prod_{j \neq i} \frac{\lambda_j - t}{\lambda_j - \lambda_i}$$

We observe that $\phi_n(t)$ is a polynomial in t of degree ν, where $\nu \leq (n-1)$.

If $\nu > 0$, the equation $\phi_n(t) = 1$ can have at most $\nu < n$ distinct real solutions.

The hypoexponential distribution (cont.)

However, for $k = 1, \ldots, n$ we must have:

$$\prod_{j \neq i} \frac{\lambda_j - \lambda_k}{\lambda_j - \lambda_i} = 0, \quad \text{if } k \neq i,$$

$$\prod_{j \neq i} \frac{\lambda_j - \lambda_k}{\lambda_j - \lambda_i} = 1, \quad \text{if } k = i.$$

Hence, we get that:

$$\phi_n(\lambda_k) = \sum_{i=1}^{n} \prod_{j \neq i} \frac{\lambda_j - \lambda_k}{\lambda_j - \lambda_i} = 1, \quad k = 1, \ldots, n.$$

Since we have assumed that $\lambda_1, \ldots, \lambda_n$ are distinct, the equation $\phi_n(t) = 1$ has n distinct real solutions, which implies that $\nu = 0$, i.e., that $\phi_n(t) \equiv 1$.

The hypoexponential distribution (cont.)

Thus, we have shown that the moment generating function of Z is simply:

$$M_Z(t) = \prod_{i=1}^{n} \frac{\lambda_i}{\lambda_i - t} = M_S(t).$$

Since the moment generating function (when it exists) uniquely determines the distribution, this implies that Z has the distribution of a sum of n independent, exponentially distributed variables with distinct rates.

NOTE: Since $\phi_n(t) \equiv 1$, it follows that we also have:

$$\sum_{i=1}^{n} C_{i,n} = \sum_{i=1}^{n} \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i} = \phi_n(0) = 1.$$

Hence, we also get that:

$$\int_0^{\infty} f_Z(z) dz = \sum_{i=1}^{n} C_{i,n} \int_0^{\infty} \lambda_i e^{-\lambda_i z} dz = \sum_{i=1}^{n} C_{i,n} = 1$$

The hypoexponential distribution (cont.)

NOTE: Both the hyperexponential distribution and the hypoexponential distribution have probability densities of the form:

$$f(x) = \sum_{i=1}^{n} a_i e^{-\lambda_i x}, \quad x \geq 0,$$

where:

$$\sum_{i=1}^{n} a_i = 1.$$

For the hyperexponential distribution the parameters $a_1, \ldots, a_n$ are all numbers between 0 and 1. Hence, these parameters can be interpreted as probabilities.

For the hypoexponential distribution the parameters $a_1, \ldots, a_n$ will in general be both less than 0 and greater than 1. Thus, in this case the parameters can not be interpreted as probabilities.

Chapter 5.2.5 The Dirichlet Distribution

Let $X_1, \ldots, X_n$ be independent and $X_i \sim gamma(\alpha_i, \lambda)$, $i = 1, \ldots, n$. Then we have shown that:

$$S = \sum_{i=1}^{n} X_i \sim gamma(\sum_{i=1}^{n} \alpha_i, \lambda).$$

We now consider the random vector:

$$(Y_1, \ldots, Y_{n-1}) = (\frac{X_1}{S}, \ldots, \frac{X_{n-1}}{S}).$$

In order to find the joint probability density of $(Y_1, \ldots, Y_{n-1})$, we first derive the joint probability density of $(Y_1, \ldots, Y_{n-1}, S)$, noting that:

$$X_i = S \cdot Y_i, \quad i = 1, \ldots, (n-1), \qquad X_n = S - \sum_{i=1}^{n-1} S \cdot Y_i = S \cdot (1 - \sum_{i=1}^{n-1} Y_i).$$

The Dirichlet Distribution (cont.)

Hence, the Jacobi matrix of this transformation is:

$$J = \begin{bmatrix} s & 0 & \dots & 0 & y_1 \\ 0 & s & \dots & 0 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & s & y_{n-1} \\ -s & -s & \dots & -s & 1 - \sum_{i=1}^{n-1} y_i \end{bmatrix}$$

Moreover, it is easy to verify that the determinant of this matrix is $|J| = s^{n-1}$.

The joint probability density of $(Y_1, \dots, Y_{n-1}, S)$ is then obtained by inserting the expressions for $x_1, \dots, x_n$ into the joint probability density of $X_1, \dots, X_n$:

$$f(y_1, \dots, y_{n-1}, s) = \prod_{i=1}^{n-1} \frac{\lambda^{\alpha_i}}{\Gamma(\alpha_i)} (sy_i)^{\alpha_i - 1} e^{-\lambda s y_i}$$

$$\cdot \frac{\lambda^{\alpha_n}}{\Gamma(\alpha_n)} (s(1 - \sum_{i=1}^{n-1} y_i))^{\alpha_n - 1} e^{-\lambda s(1 - \sum_{i=1}^{n-1} y_i)} \cdot |s^{n-1}|$$

The Dirichlet Distribution (cont.)

$$f(y_1, \ldots, y_{n-1}, s) = \prod_{i=1}^{n-1} \frac{\lambda^{\alpha_i}}{\Gamma(\alpha_i)} (sy_i)^{\alpha_i - 1} e^{-\lambda s y_i}$$

$$\cdot \frac{\lambda^{\alpha_n}}{\Gamma(\alpha_n)} (s(1 - \sum_{i=1}^{n-1} y_i))^{\alpha_n - 1} e^{-\lambda s(1 - \sum_{i=1}^{n-1} y_i)} \cdot |s^{n-1}|$$

$$= \frac{\Gamma(\alpha_1 + \cdots + \alpha_n)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)} y_1^{\alpha_1 - 1} \cdots y_{n-1}^{\alpha_{n-1} - 1} (1 - \sum_{i=1}^{n-1} y_i)^{\alpha_n - 1}$$

$$\cdot \frac{\lambda^{\alpha_1 + \cdots + \alpha_n}}{\Gamma(\alpha_1 + \cdots + \alpha_n)} s^{\alpha_1 + \cdots + \alpha_n - 1} e^{-\lambda s}.$$

The Dirichlet Distribution (cont.)

The joint probability density of $(Y_1, \ldots, Y_{n-1})$ is obtained by integrating with respect to s:

$$f(y_1,\ldots,y_{n-1}) = \int_0^\infty f(y_1,\ldots,y_{n-1},s)ds$$

$$= \frac{\Gamma(\alpha_1+\cdots+\alpha_n)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_n)} y_1^{\alpha_1-1}\cdots y_{n-1}^{\alpha_{n-1}-1}(1-\sum_{i=1}^{n-1} y_i)^{\alpha_n-1}$$

$$\cdot \int_0^\infty \frac{\lambda^{\alpha_1+\cdots+\alpha_n}}{\Gamma(\alpha_1+\cdots+\alpha_n)} s^{\alpha_1+\cdots+\alpha_n-1}e^{-\lambda s}ds$$

$$= \frac{\Gamma(\alpha_1+\cdots+\alpha_n)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_n)} y_1^{\alpha_1-1}\cdots y_{n-1}^{\alpha_{n-1}-1}(1-\sum_{i=1}^{n-1} y_i)^{\alpha_n-1}.$$

The Dirichlet Distribution (cont.)

NOTE: The joint probability density of $(Y_1, \ldots, Y_{n-1})$ and S can be factorised as follows:

$$
\begin{aligned}
f(&y_1, \ldots, y_{n-1}, s) \\
&= \frac{\Gamma(\alpha_1 + \cdots + \alpha_n)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_n)} y_1^{\alpha_1 - 1} \cdots y_{n-1}^{\alpha_{n-1}-1} (1 - \sum_{i=1}^{n-1} y_i)^{\alpha_n - 1} \\
&\cdot \frac{\lambda^{\alpha_1 + \cdots + \alpha_n}}{\Gamma(\alpha_1 + \cdots + \alpha_n)} s^{\alpha_1 + \cdots + \alpha_n - 1} e^{-\lambda s} \\
&= f(y_1, \ldots, y_{n-1}) \cdot f(s).
\end{aligned}
$$

Hence, it follows that the vector $(Y_1, \ldots, Y_{n-1})$ and S are independent of each other.

The Dirichlet Distribution (cont.)

The joint distribution of $(Y_1, \ldots, Y_{n-1})$ is called a Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_n$, and we write:

$$(Y_1, \ldots, Y_{n-1}) \sim \textit{Dirichlet}(\alpha_1, \ldots, \alpha_n)$$

NOTE: Since $Y_i = X_i/S$, it follows that:

$$\sum_{i=1}^{n-1} Y_i = \sum_{i=1}^{n-1} X_i/S = \frac{S - X_n}{S} \in (0, 1).$$

Thus, $f(y_1, \ldots, y_{n-1}) > 0$ only when $0 < y_1 + \cdots + y_{n-1} < 1$.

In the special case where $\alpha_1 = \cdots = \alpha_n = 1$, we get:

$$f(y_1, \ldots, y_{n-1}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_n)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)} y_1^{\alpha_1 - 1} \cdots y_{n-1}^{\alpha_{n-1} - 1} (1 - \sum_{i=1}^{n-1} y_i)^{\alpha_n - 1} = (n-1)!$$

The Dirichlet Distribution (cont.)

In the special case where $n = 2$, i.e., when $(Y_1, \ldots, Y_{n-1}) = Y_1$, we get:

$$f(y_1) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 - 1}(1 - y_1)^{\alpha_2 - 1}, \quad 0 < y_1 < 1.$$

This distribution is called a Beta distribution with parameters α_1, α_2.

In the case where $\alpha_1 = \alpha_2 = 1$, we get:

$$f(y_1) = \frac{\Gamma(1 + 1)}{\Gamma(1)\Gamma(1)} y_1^{1-1}(1 - y_1)^{1-1}$$

$$= 1, \quad 0 < y_1 < 1.$$

Thus, in this case $Y_1 \sim \mathit{uniform}(0, 1)$.

Chapter 5.3.1 Counting Processes

A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ represents the total number of events that occur by time $t \geq 0$.

EXAMPLES:

$N(t) =$ Number of persons arriving at a store up to and including time $t \geq 0$

$N(t) =$ Number of children being born up to and including time $t \geq 0$

$N(t) =$ Number of reported accidents up to and including time $t \geq 0$

Counting Processes (cont.)

Properties of a counting process:

- $N(t) \in \{0, 1, 2, \ldots\}$ for all $t \geq 0$.
- $N(t)$ is non-decreasing in t. That is, if $s < t$, then $N(s) \leq N(t)$.
- If $s < t$, then $N(t) - N(s)$ equals the number of events in $(s, t]$.

A counting process is said to have independent increments if $N(t_1) - N(s_1)$ and $N(t_2) - N(s_2)$ are independent for all pairs of non-overlapping intervals $(s_1, t_1]$ and $(s_2, t_2]$.

A counting process is said to have stationary increments if the distribution of $N(s + t) - N(s)$ has the same distribution for all $s \geq 0$.

The $o(h)$-notation

Definition

The function $f(\cdot)$ is said to be $o(h)$ if:

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

EXAMPLES:

(a) The function $f(x) = x^2$ is $o(h)$ since:

$$\lim_{h \to 0} \frac{f(h)}{h} = \lim_{h \to 0} \frac{h^2}{h} = \lim_{h \to 0} h = 0.$$

(b) The function $f(x) = x$ is not $o(h)$ since:

$$\lim_{h \to 0} \frac{f(h)}{h} = \lim_{h \to 0} \frac{h}{h} = \lim_{h \to 0} 1 = 1.$$

The $o(h)$-notation (cont.)

(c) If $f(\cdot)$ is $o(h)$ and $g(\cdot)$ is $o(h)$, then so is $f(\cdot) + g(\cdot)$.

$$\lim_{h \to 0} \frac{f(h) + g(h)}{h} = \lim_{h \to 0} \frac{f(h)}{h} + \lim_{h \to 0} \frac{g(h)}{h} = 0 + 0 = 0.$$

(d) If $f(\cdot)$ is $o(h)$, then so is $c \cdot f(\cdot)$.

$$\lim_{h \to 0} \frac{c \cdot f(h)}{h} = c \cdot \lim_{h \to 0} \frac{f(h)}{h} = c \cdot 0 = 0.$$

(e) Any finite linear combination of functions, each of which is $o(h)$, is $o(h)$.

Chapter 5.3.2 Definition of the Poisson Process

Definition (5.2)

The counting process $\{N(t), t \geq 0\}$ *is said to be a Poisson process with rate* $\lambda > 0$ *if the following four axioms hold:*

$$\begin{array}{ll} (i) & N(0) = 0 \\ (ii) & \{N(t), t \geq 0\} \text{ has independent increments.} \\ (iii) & P(N(t+h) - N(t) = 1) = \lambda h + o(h) \\ (iv) & P(N(t+h) - N(t) \geq 2) = o(h) \end{array}$$

Properties of the Poisson Process

Lemma (5.1)

Assume that $\{N(t), t \geq 0\}$ *is a Poisson process with rate* $\lambda > 0$*. Then let* $s > 0$*, and define:*

$$N_s(t) = N(s+t) - N(s), \qquad \text{for all } t \geq 0.$$

Then $\{N_s(t), t \geq 0\}$ *is a Poisson process with rate* λ *as well.*

PROOF: We prove the lemma by verifying that $\{N_s(t), t \geq 0\}$ satisfies the axioms given in Definition 5.2:

(i) $N_s(0) = N(s+0) - N(s) = 0$

(ii) $\{N_s(t), t \geq 0\}$ has independent increments since $\{N(t), t \geq 0\}$ has independent increments.

(iii) $P(N_s(t+h) - N_s(t) = 1) = P(N(s+t+h) - N(s+t) = 1) = \lambda h + o(h)$

(iv) $P(N_s(t+h) - N_s(t) \geq 2) = P(N(s+t+h) - N(s+t) \geq 2) = o(h)$

Properties of the Poisson Process (cont.)

Lemma (5.2)

If T_1 is the time of the first event of a Poisson process $\{N(t), t \geq 0\}$ with rate $\lambda > 0$, then:

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}, \quad t \geq 0.$$

That is $T_1 \sim exp(\lambda)$.

PROOF: Let $P_0(t) = P(N(t) = 0)$. Then:

$$\begin{aligned} P_0(t+h) &= P(N(t+h) = 0) \\ &= P(N(t) = 0 \cap N(t+h) - N(t) = 0) \\ &= P(N(t) = 0) \cdot P(N(t+h) - N(t) = 0) \qquad \text{by Axiom (ii)} \\ &= P_0(t)(1 - \lambda h + o(h)) \qquad \text{by Axiom (iii) and (iv)} \end{aligned}$$

Properties of the Poisson Process (cont.)

Hence, we get:

$$P_0(t+h) - P_0(t) = -\lambda h P_0(t) + o(h)P_0(t)$$

Dividing by h and letting $h \to 0$ gives that:

$$P_0'(t) = \lim_{h\to 0} \frac{P_0(t+h) - P_0(t)}{h} = \lim_{h\to 0}[-\lambda P_0(t) + \frac{o(h)P_0(t)}{h}] = -\lambda P_0(t).$$

Alternatively, this can be expressed as:

$$\frac{P_0'(t)}{P_0(t)} = -\lambda$$

Integrating both sides of this equation yields:

$$\log(P_0(t)) = -\lambda t + C.$$

Since $P_0(0) = P(N(0) = 0) = 1$ it follows that $C = 0$, and hence, we get:

$$P(T_1 > t) = P(N(t) = 0) = P_0(t) = e^{-\lambda t}, \quad t \geq 0 \qquad \blacksquare$$

Properties of the Poisson Process (cont.)

Proposition (5.4)

Let $T_1, T_2, \ldots$ be the sequence of interarrival times of a Poisson process $\{N(t), t \geq 0\}$ with rate $\lambda > 0$. That is, T_1 is the time of the first event, and

$$T_n = \text{The time between the } (n-1)\text{st and } n\text{th event}, \quad n = 2, 3, \ldots.$$

Then $T_1, T_2, \ldots$ are independent and identically distributed exponential random variables with rate λ.

PROOF: We have already shown that $T_1 \sim exp(\lambda)$. For T_2 we get:

$$\begin{aligned} P(T_2 > t | T_1 = s) &= P(N(s+t) - N(s) = 0 | T_1 = s) \\ &= P(N(s+t) - N(s) = 0) \qquad \text{by Axiom (ii)} \\ &= P(N_s(t) = 0) = e^{-\lambda t} \qquad \text{by Lemma 5.1 and Lemma 5.2.} \end{aligned}$$

The result follows by repeating this argument for $T_3, T_4, \ldots$ ■