Exponential Random Variables

In practice, exponential random variables model wait time until the next event from a Poisson random variable occur: If X is a Poisson with parameter λ, then the time between successive events is exponential with parameter λ.

Example

Suppose that customers arrive at a restaurant's drive-through window at a rate according to a Poisson distribution with parameter $\lambda = 100$ customers per hour. What is the expected value and standard deviation of time between customers? What is the probability of waiting at most 3 minutes for a customer? What is the probability of waiting more than 1 minute for a customer?

Since $\lambda = 100$ customers per hour, this is the same as $\lambda = \frac{5}{3}$ customer per minute

$$E(X) = \frac{1}{\lambda} = \frac{1}{\frac{5}{3}} = \frac{3}{5} = 0.6$$

$$\sigma(X) = \sqrt{\frac{1}{\lambda^2}} = \sqrt{\frac{1}{\left(\frac{5}{3}\right)^2}} = \frac{1}{\frac{5}{3}} = \frac{3}{5}$$

$$P(X \le 3) = \int_0^3 \frac{5}{3} e^{-\frac{5}{3}x}\, dx = -e^{-\frac{5}{3}x}\Big|_0^3 = -e^{-5} + 1 \approx 0.9933$$

$$\begin{aligned}\Pr(X > 1) &= 1 - \Pr(X \le 1) = 1 - \int_0^1 \frac{5}{3} e^{-\frac{5}{3}x}\, dx \\ &= 1 - \left(-e^{-\frac{5}{3}x}\right)\Big|_0^1 = 1 - \left(-e^{-\frac{5}{3}} + 1\right) = e^{-\frac{5}{3}} \approx 0.1889\end{aligned}$$

A random variable is memory-less if

$$\Pr(X > s + t \mid X > t) = \Pr(X > s) \quad \text{for all } s, t \ge 0$$

If X is the waiting time in minutes, this means that the probability of waiting at least $s + t$ minutes given that we have already waiting at least t minutes, is the same as the probability of waiting at least s minutes; the variable does not "remember" waiting those first t minutes. Exponential random variables X are memory-less since:

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

The exponential distribution is the only distribution with the memory-less property

Example

Suppose that at the previous drive-thru, it's been 6 minutes since the last customer, what is the probability that no customer arrives in the next minute? What if it's been an hour since the last customer?

$$\Pr(X > 6 + 1 \mid X > 6) = \Pr(X > 1) \approx 0.1889$$
$$\Pr(X > 60 + 1 \mid X > 60) = \Pr(X > 1) \approx 0.1889$$

Normal Random Variable

A random variable X is a normal random variable with parameters μ and σ^2 if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R}$$

As you know from STAT 231, Normal distributions are bell shaped without outliers. They model many natural phenomena such as weights of animals, velocities of gas molecules, blood pressures of adults, measurement error in an experiment, etc.

The 68-95-99.7 Rule:

- $\Pr(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$
- $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$
- $\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$

If X_i is a family of independent normal random variables with parameters μ_i and σ_i^2, then $Y = \sum c_i X_i$ is also normal with mean $\sum c_i \mu_i$ and variance $\sum c_i^2 \sigma_i^2$.

Example

Three species have weights (lbs) that are normally distributed:

- Species 1 has weight X_1 with mean and variance $\mu_1 = 10$ and $\sigma_1^2 = 3$
- Species 2 has weight X_3 with mean and variance $\mu_2 = 8$ and $\sigma_2^2 = 4$
- Species 3 has weight X_3 with mean and variance $\mu_3 = 11$ and $\sigma_3^2 = 2$

Suppose that one animal of each species is randomly chosen for transport. The transport fee is \$5 per lb for Species 1 and 2, and \$10 per lb for Species 3. What is the expected value and variance of the transport cost for these animals?

Let Y be the total transport cost: $Y = 5X_1 + 5X_2 + 10X_3$.

$$\begin{aligned} E\left(Y\right) &= E\left(5X_1 + 5X_2 + 10X_3\right) \\ &= 5 \cdot E\left(X_1\right) + 5 \cdot E\left(X_2\right) + 10 \cdot E\left(X_3\right) \\ &= 5 \cdot 10 + 5 \cdot 8 + 10 \cdot 11 \\ &= 200 \end{aligned}$$

$$\begin{aligned} V\left(Y\right) &= V\left(5X_1 + 5X_2 + 10X_3\right) \\ &= 5 \cdot V\left(X_1\right) + 5 \cdot V\left(X_2\right) + 10 \cdot V\left(X_3\right) \\ &= 25 \cdot 3 + 25 \cdot 4 + 100 \cdot 2 \\ &= 375 \end{aligned}$$

Standard Normal Distribution

A normal random variable is called standard if $\mu = 0$ and $\sigma^2 = 1$, typically denoted by Z. The cdf of the standard normal distribution is usually denoted by Φ:

$$\Phi\left(z\right) = \Pr\left(Z \le z\right) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

In general, $\Phi\left(z\right)$ is not easy to compute, but it's values are in the Z-Table. Also, $\Phi\left(-z\right) = 1 - \Phi\left(z\right)$, by symmetry. If X is a normal random variable with mean and variance μ and σ^2, then

$$Z = \frac{X - \mu}{\sigma}$$

is a standard normal variable and:

$$\Pr\left(a \le X \le b\right) = \Pr\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Example

Suppose that X is a normal random variable with mean $\mu = 2$ and variance $\sigma^2 = 25$. Calculate $\Pr(1 \leq X \leq 4)$, $\Pr(X > 0)$, $\Pr\left((X-2)^2 > 5\right)$.

$$\begin{aligned}
\Pr(1 \leq X \leq 4) &= \Pr\left(-\frac{1}{5} \leq Z \leq \frac{2}{5}\right) \\
&= \Phi\left(\frac{2}{5}\right) - \Phi\left(-\frac{1}{5}\right) \\
&= 0.6554 - 0.4207 = 0.2347
\end{aligned}$$

$$\begin{aligned}
\Pr(X > 0) &= \Pr(Z > -2.5) \\
&= 1 - \Pr(Z < -2.5) \\
&= 1 - \Phi\left(-\frac{2}{5}\right) \\
&= 1 - 0.3446 \\
&= 0.6554
\end{aligned}$$

$$\begin{aligned}
\Pr\left((X-2)^2 > 5\right) &= \Pr\left(|X-2| > \sqrt{5}\right) \\
&= \Pr\left(\left((X-2) > \sqrt{5}\right) \cup \left((X-2) < -\sqrt{5}\right)\right) \\
&= \Pr\left(\frac{X-2}{5} > \frac{\sqrt{5}}{5}\right) + \Pr\left(\frac{X-2}{5} < -\frac{\sqrt{5}}{5}\right) \\
&= \Pr(Z > 0.4472) + \Pr(Z < -0.4472) \\
&= (1 - \Phi(0.4472)) + (\Phi(-0.4472)) \\
&= 0.3274 + 0.3274 \\
&= 0.6548
\end{aligned}$$

Normal Approximation to Binomial

If X is a binomial random variable with parameters (n, α). As $n\alpha \to \infty$, X is approximately normal with $\mu = n\alpha$ and $\sigma^2 = n\alpha(1-\alpha)$. This is a good enough approximation if $n\alpha \geq 10$ and $n(1-\alpha) \geq 10$.

Example

If 10% of people own red cars, what is the probability that fewer than 100 in a random sample of 818 people own a red car?

Let X be the number of people in the sample who own a red car. Then, $X \sim \text{Bin}\left(818, 0.1\right)$, so $n = 818, \alpha = 0.1$. The exact probability (using R) is:

$$\Pr\left(X < 100\right) = \sum_{x=0}^{99} \binom{818}{x} \left(0.1\right)^x \left(0.9\right)^{818-x} = 0.9782$$

Since $n\alpha = \left(818\right)\left(0.1\right) \geq 10$ and $n\left(1-\alpha\right) = \left(818\right)\left(0.9\right) \geq 10$, then we can approximate this using a normal distribution.

$$\mu = \left(818\right)\left(0.1\right) = 81.8 \qquad \sigma^2 = n\alpha\left(1-\alpha\right) = \left(818\right)\left(0.1\right)\left(0.9\right) = 73.62$$

So, $X \sim N\left(\mu = 81.8, \sigma^2 = 73.62\right)$. This also means $\sigma = \sqrt{73.62} = 8.58$. So,

$$\begin{aligned}\Pr\left(X < 100\right) &= \Pr\left(\frac{X - 81.8}{8.58} < \frac{100 - 81.8}{8.58}\right) \\ &= \Pr\left(Z < 2.12\right) \\ &= 0.9830\end{aligned}$$

The Lognormal Distribution

Suppose that X is a normal random variable with mean μ and variance σ^2. Then $Y = e^X$ is a lognormal random variable with parameters μ and σ^2. (We call it lognormal because $\ln Y$ is a normal random variable). What is the pdf of Y?

We know that $X \sim N\left(\mu, \sigma^2\right)$ and $Y = e^X$. We want $Y = e^X = g\left(X\right) \Leftrightarrow X = \ln\left(Y\right) = g^{-1}\left(Y\right)$, since we know $Y > 0$. Then use change of variables:

$$f_Y\left(y\right) = f_X\left(g^{-1}\left(x\right)\right) \cdot \left|\frac{d}{dy} g^{-1}\left(y\right)\right|$$

Here, $f_X\left(x\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(x-\mu\right)^2}{2\sigma^2}}$, and $g^{-1}\left(y\right) = \ln\left(y\right)$, so $\frac{d}{dy} g^{-1}\left(y\right) = \frac{1}{y}$. Substituting these in gives us:

$$f_Y\left(y\right) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{\left(\ln\left(y\right)-\mu\right)^2}{2\sigma^2}} \qquad \text{for } y > 0$$

So, the pdf of the lognormal distribution is

$$f_Y\left(y\right) = \begin{cases} f_Y\left(y\right) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{\left(\ln\left(y\right)-\mu\right)^2}{2\sigma^2}} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$