Probability III:
Some More (mostly continuous) Distrubutions

Elie Gurarie

Biol 799 - Lecture 5b

January 5, 2017

The Volcano: Part I

UW vulcanologist says: “Mt. Rainier will definitely erupt in the next 100 years, but it could happen any time.”

Continuous random variable

- Eruption time X is a **continuous random variable**.
- If eruption is equally likely to occur at any moment between now and the 100 year interval, then:
  - $P(X < 50 \text{ years}) = 0.5$
  - $P(X < 10 \text{ years}) = 0.1$
  - $P(X > 80 \text{ years}) = 0.2$
  - $P(50 < X < 80 \text{ years}) = 0.3$
- BUT
  - $P(X = 50 \text{ years}) = P(X = 10 \text{ years}) = P(X = 86.593832715 \text{ years}) = 0$
- How do you calculate these numbers?

Continuous random variable

- Eruption time X is a **continuous random variable**.
- If eruption is equally likely to occur at any moment between now and the 100 year interval, then:
  - $P(X < 50 \text{ years}) = 0.5$
  - $P(X < 10 \text{ years}) = 0.1$
  - $P(X > 80 \text{ years}) = 0.2$
  - $P(50 < X < 80 \text{ years}) = 0.3$
- BUT
  - $P(X = 50 \text{ years}) = P(X = 10 \text{ years}) = P(X = 86.593832715 \text{ years}) = 0$
- How do you calculate these numbers?

Continuous random variable

- Eruption time X is a **continuous random variable**.
- If eruption is equally likely to occur at any moment between now and the 100 year interval, then:
  - $P(X < 50 \text{ years}) = 0.5$
  - $P(X < 10 \text{ years}) = 0.1$
  - $P(X > 80 \text{ years}) = 0.2$
  - $P(50 < X < 80 \text{ years}) = 0.3$
- BUT
  - $P(X = 50 \text{ years}) = P(X = 10 \text{ years}) = P(X = 86.593832715 \text{ years}) = 0$
- How do you calculate these numbers?

Probability density function $f(x)$ (pdf)

- $f(x) \geq 0$ for all values of x.
- Total area under the curve of $f(x)$ is 1.

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

- $P(a < X < b) = \int_a^b f(x)\, dx$

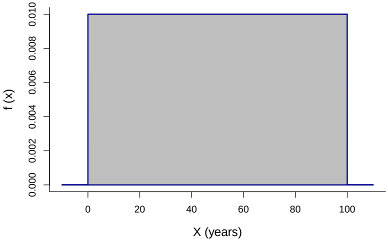
Uniform random variable

Models an event that can happen with equal probability at *any* moment between time α and time β.

- $X \sim \text{Unif}(\alpha, \beta)$
- $f(x) = \begin{cases} \frac{1}{\beta-\alpha} \text{ where } \alpha \leq x \leq \beta \\ 0 \text{ otherwise.} \end{cases}$
- $$\begin{aligned} P(a < X < b) &= \int_a^b \frac{1}{\beta - \alpha} \, dx \\ &= \frac{b-a}{\beta - \alpha} \end{aligned}$$
- Known as: **continuous uniform distribution**.

Uniform distribution

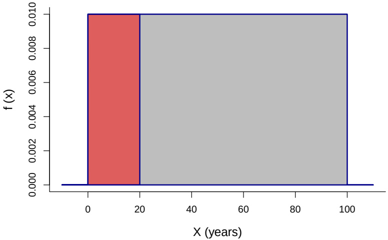
- $X \sim \text{Unif}(0, 100)$
- $f(x) = \begin{cases} \frac{1}{100} \text{ where } 0 \leq x \leq 100 \\ 0 \text{ else} \end{cases}$



Uniform distribution

- $X \sim \text{Unif}(0, 100)$
- $f(x) = \begin{cases} \frac{1}{100} \text{ where } 0 \leq x \leq 100 \\ 0 \text{ else} \end{cases}$

$$\begin{aligned} P(X < 20) &= \int_{-\infty}^{20} \frac{1}{100} dx \\ &= \frac{20}{100} - \frac{0}{100} \\ &= 20 \end{aligned}$$

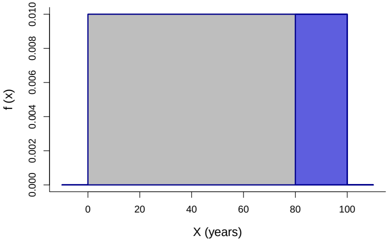



```
punif(20,0,100)
```

Uniform distribution

- $X \sim \text{Unif}(0, 100)$
- $f(x) = \begin{cases} \frac{1}{100} \text{ where } 0 \leq x \leq 100 \\ 0 \text{ else} \end{cases}$

$$\begin{aligned} P(X > 80) &= \int_{80}^{\infty} \frac{1}{100} dx \\ &= \frac{100}{100} - \frac{80}{100} \\ &= 20 \end{aligned}$$



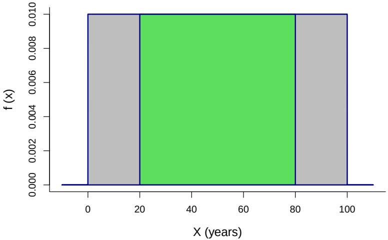

```
1 - punif(80,0,100)
```

Uniform distribution

- $X \sim \text{Unif}(0, 100)$
- $f(x) = \begin{cases} \frac{1}{100} \text{ where } 0 \leq x \leq 100 \\ 0 \text{ else} \end{cases}$

$$\begin{aligned} P(20 < X < 80) &= \int_{20}^{80} \frac{1}{100} dx \\ &= \frac{80}{100} - \frac{20}{100} \\ &= 60 \end{aligned}$$

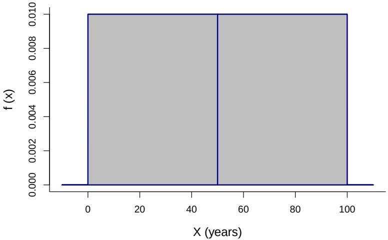
```
punif(80,0,100) - punif(20,0,100)
```



Uniform distribution

- $X \sim \text{Unif}(0, 100)$
- $f(x) = \begin{cases} \frac{1}{100} \text{ where } 0 \leq x \leq 100 \\ 0 \text{ else} \end{cases}$

$$
\begin{aligned}
P(X = 50) &= \int_{50}^{50} \frac{1}{100} dx \\
&= \frac{50}{100} - \frac{50}{100} \\
&= 0
\end{aligned}
$$




```
punif(20,0,100)
```

Expected value and variance

- Refresher: Let X be a discrete random variable with possible values $x_1, x_2, x_3, ..., x_n$ and probability mass function $f(x)$
- $\mathrm{E}(X) = \sum_{i=1}^{n} x_i f(x_i)$
- $\mathrm{Var}(X) = \mathrm{E}(([X - \mathrm{E}(X)]^2)) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \sum_{i=1}^{n} (x_i - \mathrm{E}(X))^2 f(x_i)$

Expectation and variance for continuous random variables

$$
\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x\, f(x)\, dx \\
\mathrm{Var}(X) &= \mathrm{E}(([X - \mathrm{E}(X)]^2)) \\
&= \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2\, f(x)\, dx \\
\mathrm{SD}(X) &= \sqrt{\mathrm{Var}(X)}
\end{aligned}
$$

Expected value and variance

- Refresher: Let X be a discrete random variable with possible values $x_1, x_2, x_3, ..., x_n$ and probability mass function $f(x)$
- $\mathrm{E}(X) = \sum_{i=1}^{n} x_i f(x_i)$
- $\mathrm{Var}(X) = \mathrm{E}(([X - \mathrm{E}(X)]^2)) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \sum_{i=1}^{n} (x_i - \mathrm{E}(X))^2 f(x_i)$

Expectation and variance for continuous random variables

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x\, f(x)\, dx \\
\mathrm{Var}(X) &= \mathrm{E}(([X - \mathrm{E}(X)]^2)) \\
&= \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2\, f(x)\, dx \\
\mathrm{SD}(X) &= \sqrt{\mathrm{Var}(X)}
\end{aligned}$$

Uniform random variable

Let $X \sim \text{Unif}(\alpha, \beta)$, $f(x) = \frac{1}{\beta-\alpha}$

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_{\alpha}^{\beta} x \frac{1}{\beta-\alpha}\, dx \\
&= \frac{1}{\beta-\alpha} \frac{x^2}{2}\bigg|_{\alpha}^{\beta} = \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} \\
&= \frac{(\beta-\alpha)(\beta+\alpha)}{2(\beta-\alpha)} = \frac{\beta+\alpha}{2}
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}(X) &= \int_{-\infty}^{\infty} (x - E(X))^2\, f(x)\, dx \\
&= \int_{\alpha}^{\beta} \left(x - \frac{\beta+\alpha}{2}\right)^2 \left(\frac{1}{\beta-\alpha}\right) dx \\
&= \cdots \\
\text{Var}(X) &= \frac{1}{12}(\beta-\alpha)^2
\end{aligned}
$$

Uniform random variable

Let $X \sim \text{Unif}(\alpha, \beta)$, $f(x) = \frac{1}{\beta-\alpha}$

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_{\alpha}^{\beta} x \frac{1}{\beta-\alpha}\, dx \\
&= \frac{1}{\beta-\alpha} \frac{x^2}{2}\Big|_{\alpha}^{\beta} = \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} \\
&= \frac{(\beta-\alpha)(\beta+\alpha)}{2(\beta-\alpha)} = \frac{\beta+\alpha}{2}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x-\mathrm{E}(X))^2\, f(x)\, dx \\
&= \int_{\alpha}^{\beta} \left(x - \frac{\beta+\alpha}{2}\right)^2 \left(\frac{1}{\beta-\alpha}\right) dx \\
&= \cdots \\
\mathrm{Var}(X) &= \frac{1}{12}(\beta-\alpha)^2
\end{aligned}$$

Uniform random variable

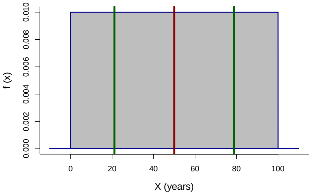
Let $X \sim \text{Unif}(\alpha, \beta)$, $f(x) = \frac{1}{\beta - \alpha}$

$$\begin{aligned}
\text{E}(X) &= \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_{\alpha}^{\beta} x \frac{1}{\beta - \alpha}\, dx \\
&= \frac{1}{\beta - \alpha} \frac{x^2}{2} \bigg|_{\alpha}^{\beta} = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\
&= \frac{(\beta - \alpha)(\beta + \alpha)}{2(\beta - \alpha)} = \frac{\beta + \alpha}{2}
\end{aligned}$$

$$\begin{aligned}
\text{Var}(X) &= \int_{-\infty}^{\infty} (x - \text{E}(X))^2\ f(x)\, dx \\
&= \int_{\alpha}^{\beta} \left(x - \frac{\beta + \alpha}{2}\right)^2 \left(\frac{1}{\beta - \alpha}\right) dx \\
&= \ldots \\
\text{Var}(X) &= \frac{1}{12} (\beta - \alpha)^2
\end{aligned}$$

Uniform random variable

- Eruption time $X \sim \text{Unif}(0, 100)$
- $\text{E}(X) = 100/2 = 50$
- $\text{Var}(X) = 100^2/12 = 833.33$
- $\text{SD}(X) = 100/\sqrt{12} = 28.87$



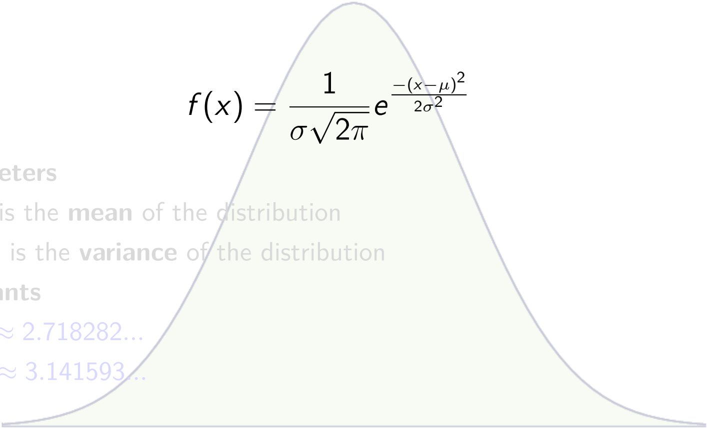
$X \sim \text{Unif}(\alpha, \beta)$

Name:	Uniform distribution	Models equal probability events within a continuous range of values
Support:	x	$(-\infty, \infty)$
Parameters:	$\alpha \in (-\infty, \beta)$ $\beta \in (\alpha, \infty)$	minimum maximum
pdf	$f(x\|\alpha, \beta) =$	$\frac{1}{\beta - \alpha}$
mean	$E(X)$	$\frac{\beta + \alpha}{2}$
variance	$\text{Var}(X)$	$\frac{1}{12}(\beta - \alpha)^2$

The Volcano Part II

Japanese vulcanologist says: "Mt. Fuji is such a beautiful volcano, it will have the most beautiful distribution, and erupt with a normal distribution with mean 50 and standard deviation 20 years."

Normal distribution: The formula



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

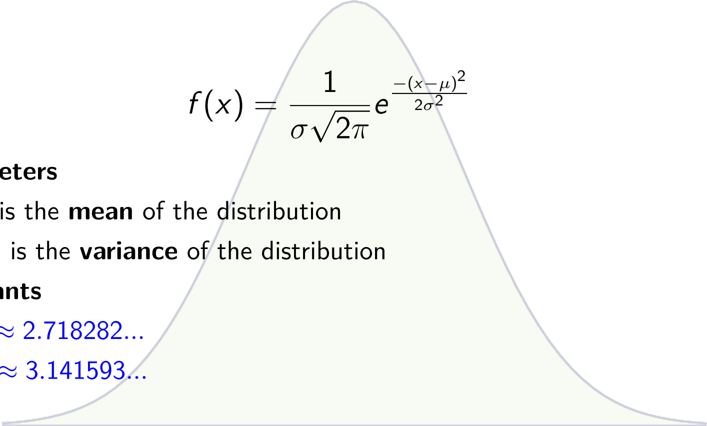
Parameters

- μ is the **mean** of the distribution
- σ^2 is the **variance** of the distribution

Constants

- $e \approx 2.718282...$
- $\pi \approx 3.141593...$

Normal distribution: The formula



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

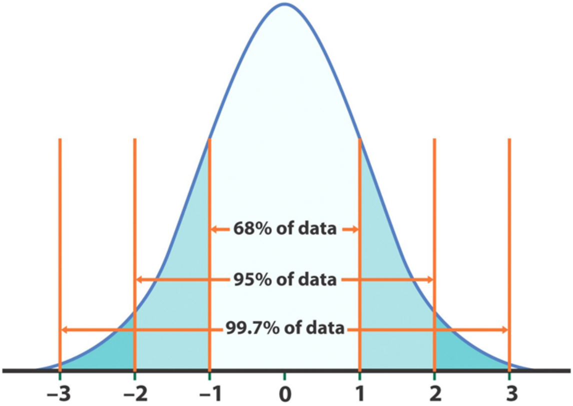
Parameters

- μ is the **mean** of the distribution
- σ^2 is the **variance** of the distribution

Constants

- $e \approx 2.718282...$
- $\pi \approx 3.141593...$

No matter what the parameters are...

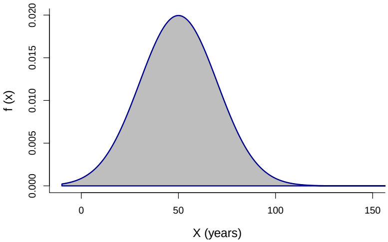


Approximately:

- 68% of the probability lies "within one standard deviations" ($\mu \pm \sigma$).
- 95% of the probability lies "within 2 standard deviations"($\mu \pm 2\sigma$).
- 99.7% of the probability lies "within 3 standard deviations"($\mu \pm 3\sigma$).

Normal distribution

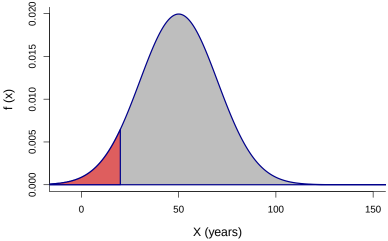
- $X \sim \mathrm{N}(\mu = 50, \sigma = 20)$
- $f(x) = \frac{1}{20\sqrt{2\pi}} e^{\frac{-(x-50)^2}{220^2}}$



Normal distribution

- $X \sim \mathrm{N}(\mu = 50, \sigma = 20)$
- $f(x) = \frac{1}{20\sqrt{2\pi}} e^{\frac{-(x-50)^2}{220^2}}$

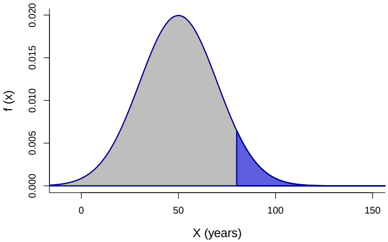
- $P(X < 20) = \int_{-\infty}^{20} f(x)dx =$ 0.067



Normal distribution

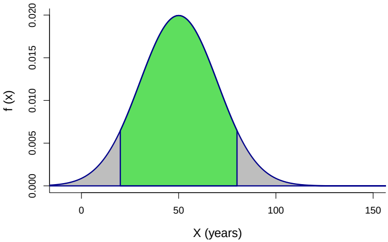
- $X \sim \mathrm{N}(\mu = 50, \sigma = 20)$
- $f(x) = \frac{1}{20\sqrt{2\pi}} e^{\frac{-(x-50)^2}{2 20^2}}$

- $P(X > 80) = \int_{80}^{\infty} f(x)dx = 0.067$



Normal distribution

- $X \sim \mathrm{N}(\mu = 50, \sigma = 20)$
- $f(x) = \frac{1}{20\sqrt{2\pi}} e^{\frac{-(x-50)^2}{220^2}}$
- $P(20 < X < 80) = 0.866$



$X \sim \text{Normal}(\mu, \sigma)$

Name:	Normal distribution	Models bell-shaped continuous variables.
Support:	x	$(-\infty, \infty)$
Parameters:	$\mu \in (-\infty, \infty)$ $\sigma \in (0, \infty)$	measure of center measure of spread
pdf	$f(x\|\mu, \sigma) =$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
mean	$E(X)$	μ
variance	$\text{Var}(X)$	σ^2

Note on the normal distribution

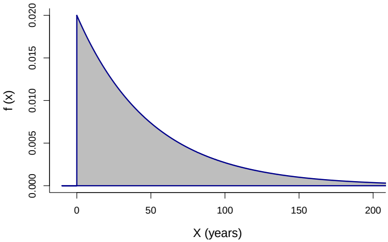
The normal distribution is a fantastic distribution for many things, but it is a lousy model for volcano eruptions!

The Volcano Part III

Russian vulcanologist says: “Mt. Avacha has been erupting at totally random times, but on average every 50 years.”

Exponential distribution

- $X \sim \text{Exp}(\gamma = 50)$
- $f(x) = \frac{1}{50} e^{-x/50}$, for $x > 0$



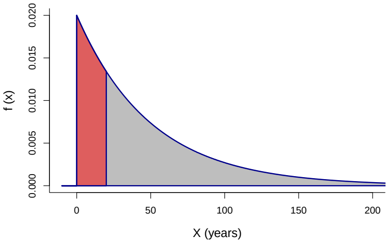
Exponential distribution

- $X \sim \text{Exp}(\gamma = 50)$
- $f(x) = \frac{1}{50} e^{-x/50}$, for $x > 0$

$P(X < 20) = \int_{-\infty}^{20} f(x)dx = 0.33$

```
pexp(20, rate=1/50)
```

Note, the use of "rate $= 1/\gamma$" instead of scale.

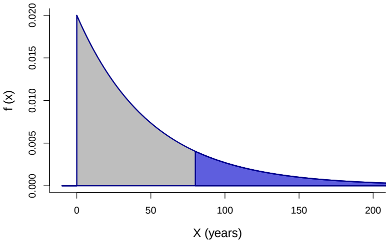


Exponential distribution

- $X \sim \text{Exp}(\gamma = 50)$
- $f(x) = \frac{1}{50} e^{-x/50}$, for $x > 0$

$P(X > 80) = \int_{80}^{\infty} f(x)dx = 0.20$

```
1-pexp(80, rate=1/50)
```

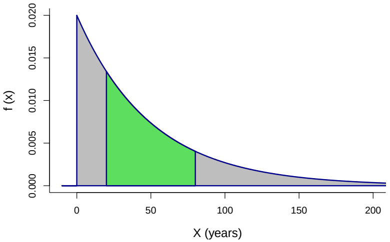


Exponential distribution

- $X \sim \text{Exp}(\gamma = 50)$
- $f(x) = \frac{1}{50} e^{-x/50}$, for $x > 0$

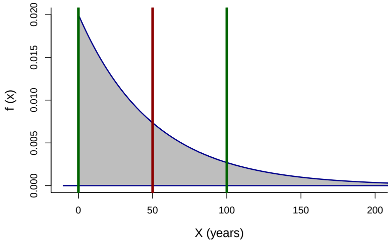
$P(20 < X < 80) = 0.47$

```
pexp(80, 1/50) - pexp(20,
1/50)
```



Exponential Distribution

- Eruption time
  $X \sim \text{Exp}(\gamma = 50)$
- $\text{E}(X) = 50$
- $\text{Var}(X) = 50^2$
- $\text{SD}(X) = 50$



Exponential random variable

Let $X \sim \mathrm{Exp}(\gamma)$, $f(x) = \frac{1}{\gamma} e^{\frac{-x}{\gamma}}$

$$\begin{aligned} \mathrm{E}(X) &= \int_0^\infty x\, f(x)\, dx = \int_0^\infty \frac{1}{\gamma} e^{\frac{-x}{\gamma}}\, dx \\ &= e^{\frac{-x}{\gamma}} (\gamma + x) \Big|_0^\infty \\ &= \gamma \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X) &= \int_0^\infty (x-\gamma)^2\, f(x)\, dx \\ &= \ldots \\ &= \gamma^2 \end{aligned}$$

$$\mathrm{SD}(X) = \gamma$$

Exponential random variable

Let $X \sim \text{Exp}(\gamma)$, $f(x) = \frac{1}{\gamma} e^{\frac{-x}{\gamma}}$

$$\begin{aligned}
\text{E}(X) &= \int_0^\infty x\, f(x)\, dx = \int_0^\infty \frac{1}{\gamma} e^{\frac{-x}{\gamma}}\, dx \\
&= e^{\frac{-x}{\gamma}} (\gamma + x) \Big|_0^\infty \\
&= \gamma
\end{aligned}$$

$$\begin{aligned}
\text{Var}(X) &= \int_0^\infty (x-\gamma)^2 f(x)\, dx \\
&= \ldots \\
&= \gamma^2
\end{aligned}$$

$$\text{SD}(X) = \gamma$$

Exponential random variable

Let $X \sim \text{Exp}(\gamma)$, $f(x) = \frac{1}{\gamma} e^{\frac{-x}{\gamma}}$

$$\begin{aligned}
\text{E}(X) &= \int_0^\infty x\, f(x)\, dx = \int_0^\infty \frac{1}{\gamma} e^{\frac{-x}{\gamma}}\, dx \\
&= e^{\frac{-x}{\gamma}} (\gamma + x) \Big|_0^\infty \\
&= \gamma
\end{aligned}$$

$$\begin{aligned}
\text{Var}(X) &= \int_0^\infty (x-\gamma)^2 f(x)\, dx \\
&= \ldots \\
&= \gamma^2
\end{aligned}$$

$$\text{SD}(X) = \gamma$$

Exponential random variable

Let $X \sim \text{Exp}(\gamma)$, $f(x) = \frac{1}{\gamma} e^{\frac{-x}{\gamma}}$

$$\begin{aligned} \text{E}(X) &= \int_0^\infty x\, f(x)\, dx = \int_0^\infty \frac{1}{\gamma} e^{\frac{-x}{\gamma}}\, dx \\ &= e^{\frac{-x}{\gamma}} (\gamma + x) \Big|_0^\infty \\ &= \gamma \end{aligned}$$

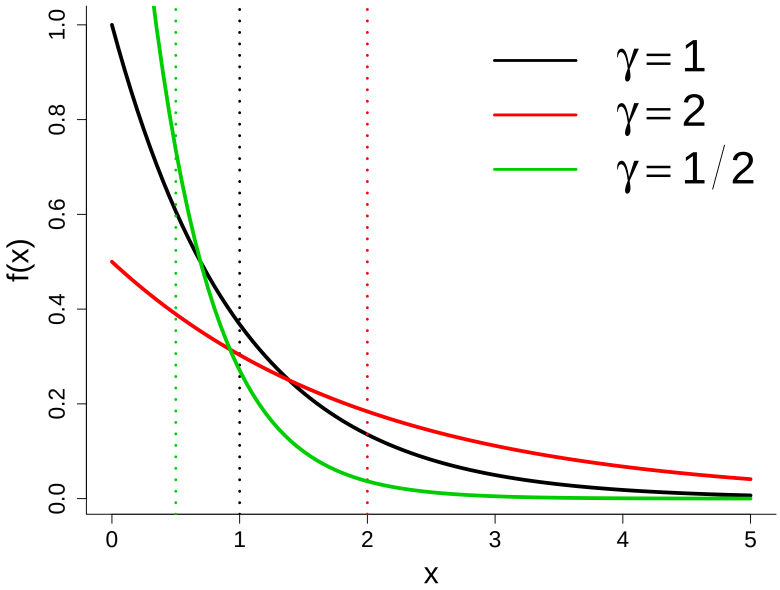
$$\begin{aligned} \text{Var}(X) &= \int_0^\infty (x - \gamma)^2\, f(x)\, dx \\ &= \ldots \\ &= \gamma^2 \end{aligned}$$

$$\text{SD}(X) = \gamma$$

$X \sim \text{Exp}(\gamma)$

Name:	Exponential distribution	Models time-independent random events
Support:	x	$(0, \infty)$
Parameters:	$\gamma \in (0, \infty)$	scale parameter
pdf	$f(x\|\mu, \sigma) =$	$\frac{1}{\gamma} e^{-\frac{x}{\gamma}}$
mean	$\text{E}(X)$	γ
variance	$\text{Var}(X)$	γ^2

Exponential Distribution: $X \sim \text{Exp}(\gamma)$



Important comments on the exponential distribution

- The exponential distribution models the waiting time to any event which can occur with equal probability in time.
- It is the continuous analogue of the geometric distribution, and is also memoryless:
  - If Avacha erupts today, the expected time to next eruption is 50 years.
  - If Avacha hasn't erupted in 50 years, the expected time to next eruption is 50 years.
  - If Avacha hasn't erupted in 500 years, the expected time to next eruption is 50 years (though you might consider updating your Exp(50) model).
- It is readily identified by the standard deviation being similar to the mean.

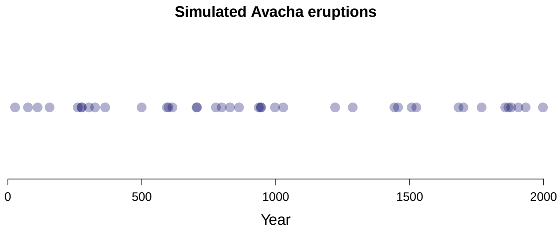
Important comments on the exponential distribution

- The exponential distribution models the waiting time to any event which can occur with equal probability in time.
- It is the continuous analogue of the geometric distribution, and is also memoryless:
  - If Avacha erupts today, the expected time to next eruption is 50 years.
  - If Avacha hasn't erupted in 50 years, the expected time to next eruption is 50 years.
  - If Avacha hasn't erupted in 500 years, the expected time to next eruption is 50 years (though you might consider updating your Exp(50) model).
- It is readily identified by the standard deviation being similar to the mean.

Important comments on the exponential distribution

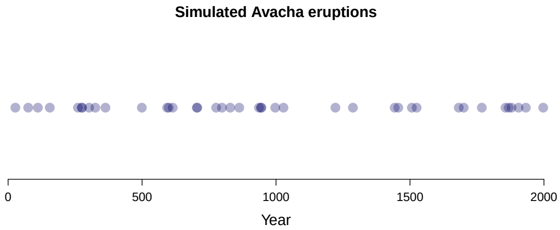
- The exponential distribution models the waiting time to any event which can occur with equal probability in time.
- It is the continuous analogue of the geometric distribution, and is also memoryless:
  - If Avacha erupts today, the expected time to next eruption is 50 years.
  - If Avacha hasn't erupted in 50 years, the expected time to next eruption is 50 years.
  - If Avacha hasn't erupted in 500 years, the expected time to next eruption is 50 years (though you might consider updating your Exp(50) model).
- It is readily identified by the standard deviation being similar to the mean.

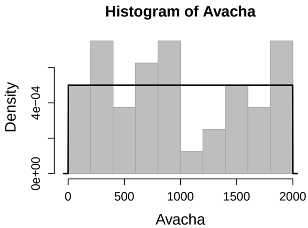
Relationship between exponential and uniform distributions




```
> Avacha <- runif(40,0,1000)
```

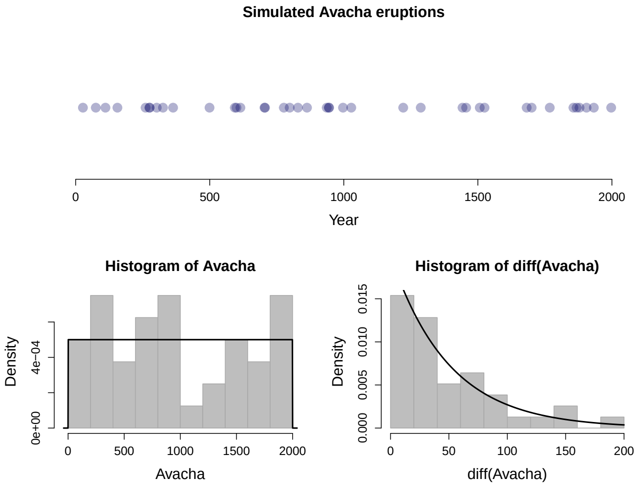
Relationship between exponential and uniform distributions





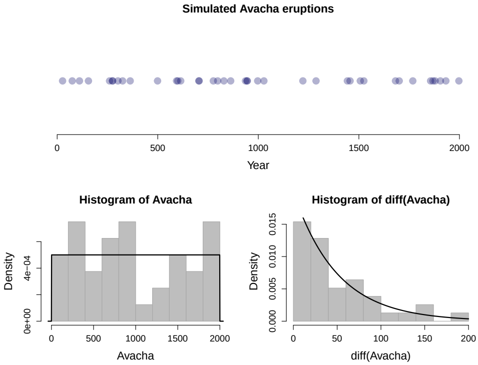

```
> Avacha <- runif(40,0,1000)
```

Relationship between exponential and uniform distributions




```
> Avacha <- runif(40,0,1000)
```

Relationship between exponential and uniform distributions

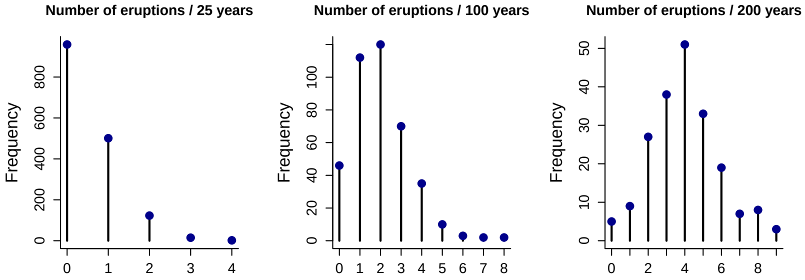


Important fact: If $X_{(1)}, X_{(2)}, X_{(3)} ... X_{(n)}$ are ordered uniform r.v.'s: $X \sim \text{Unif}(0, l)$, then:

$$W_i = X_{(i+1)} - X_{(i)} \sim \text{Exp}(\gamma = \frac{n}{l}).$$

More questions about exponential distribution

Ok, if Avacha erupts randomly with mean 50 years. How many eruptions can we *expect* in a century? 200 years? 25 years?



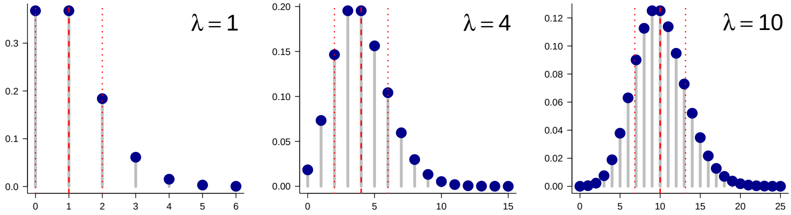
Poisson distribution

- Describes the number of times a random, independent event occurring with constant intensity in a given interval of time or space

$$P(c = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

where λ = rate or intensity of occurrence

Poisson process



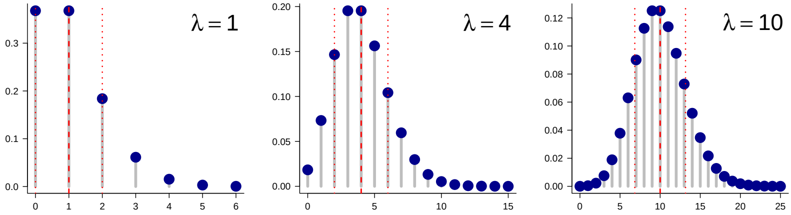
For $X \sim \text{Poisson}(\lambda)$

$$\begin{aligned} \text{E}[X] &= \lambda \\ \text{Var}[X] &= \lambda \end{aligned}$$

i.e.: **Variance = Mean = Intensity!**

This is a very useful property for determining whether the Poisson distribution is an appopriate model.

Poisson process



For $X \sim \text{Poisson}(\lambda)$

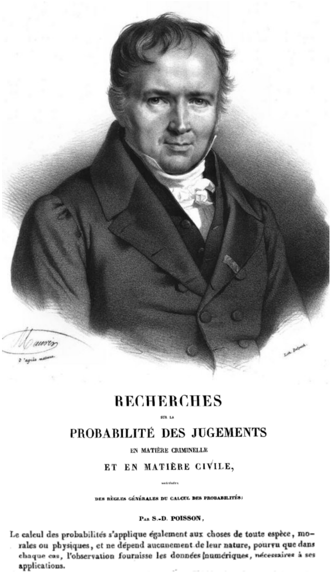
$$\begin{aligned} \text{E}[X] &= \lambda \\ \text{Var}[X] &= \lambda \end{aligned}$$

i.e.: **Variance = Mean = Intensity!**
This is a very useful property for determining whether the Poisson distribution is an appopriate model.

$X \sim \text{Poisson}(\lambda)$

Name:	Poisson distribution	Models frequency of discrete random events
Support:	x	$\{0, 1, 2, 3, ...\}$
Parameters:	$\lambda \in (0, \infty)$	intensity parameter
pdf	$P(X = k \mid \lambda) =$	$\frac{\lambda^k}{k!} e^{-\lambda}$
mean	$\text{E}(X)$	λ
variance	$\text{Var}(X)$	λ

Historical Aside



Siméon Denis Poisson (1781-1840) - French physicist and mathematician, developed the Poisson distribution to model the number of convictions in the civil courts in France, noting in 1837 that:

"The science of probability can be applied to any subject - be they moral or physical - regardless of their nature, as long as the observations provide the numerical data required for its application."

Volcano Model

Which of these distributions best describes actual eruption times?
R lab on data from Mt. Vesuvius.