The Exponential Distribution

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Earthquake

Goal: Model time until next earthquake $\tilde{t}$

Assumption: Probability of earthquake in period of length ϵ is $\lambda\epsilon$, no matter what happened before

$$\mathrm{P}\left(t \leq \tilde{t} \leq t + \epsilon \,|\, \tilde{t} > t\right) \approx \lambda\epsilon$$

Distribution of $\tilde{t}$?

Earthquake

$$S(t) := 1 - F_{\tilde{t}}(t)$$

$$\begin{aligned} \mathrm{P}\left(t < \tilde{t} \leq t + \epsilon \,\middle|\, \tilde{t} > t\right) &= \frac{\mathrm{P}\left(t < \tilde{t} \leq t + \epsilon, \tilde{t} > t\right)}{\mathrm{P}\left(\tilde{t} > t\right)} \\ &= \frac{\mathrm{P}\left(t < \tilde{t} \leq t + \epsilon\right)}{\mathrm{P}\left(\tilde{t} > t\right)} \\ &= \frac{F_{\tilde{t}}(t + \epsilon) - F_{\tilde{t}}(t)}{1 - F_{\tilde{t}}(t)} \\ &= \frac{S(t) - S(t + \epsilon)}{S(t)} \approx \lambda\epsilon \end{aligned}$$

$$-\lambda = \frac{1}{S(t)} \lim_{\epsilon \to 0} \frac{S(t + \epsilon) - S(t)}{\epsilon}$$

Earthquake

$$
\begin{aligned}
-\lambda &= \frac{1}{S(t)} \lim_{\epsilon \to 0} \frac{S(t+\epsilon) - S(t)}{\epsilon} \\
&= \frac{1}{S(t)} \frac{\mathrm{d}S(t)}{\mathrm{d}t} \\
&= \frac{\mathrm{d}\log S(t)}{\mathrm{d}t}
\end{aligned}
$$

$$
-\lambda t + c = \log S(t)
$$

$$
c' \exp(-\lambda t) = S(t) = 1 - F_{\tilde{t}}(t)
$$

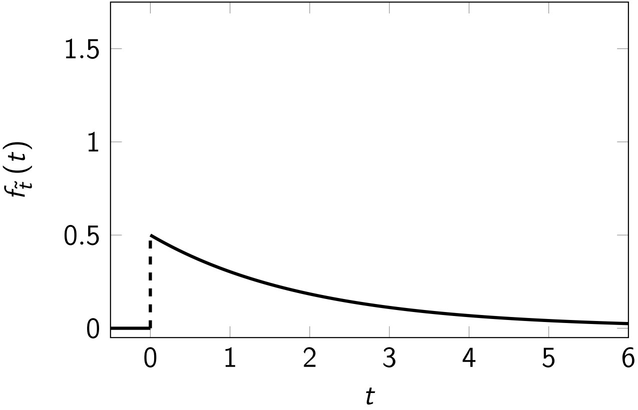
$$
\begin{aligned}
F_{\tilde{t}}(t) &= 1 - c' \exp(-\lambda t) = 1 - \exp(-\lambda t) \\
f_{\tilde{t}}(t) &= \frac{\mathrm{d}F_{\tilde{t}}(t)}{\mathrm{d}t} = \lambda \exp(-\lambda t)
\end{aligned}
$$

Exponential distribution

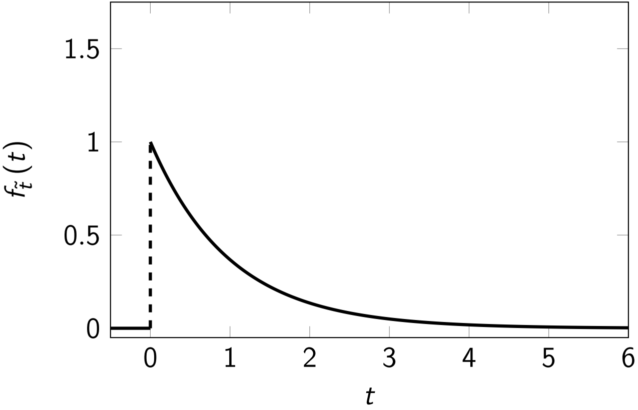
The pdf of an exponential random variable $\tilde{t}$ with parameter λ is

$$f_{\tilde{t}}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

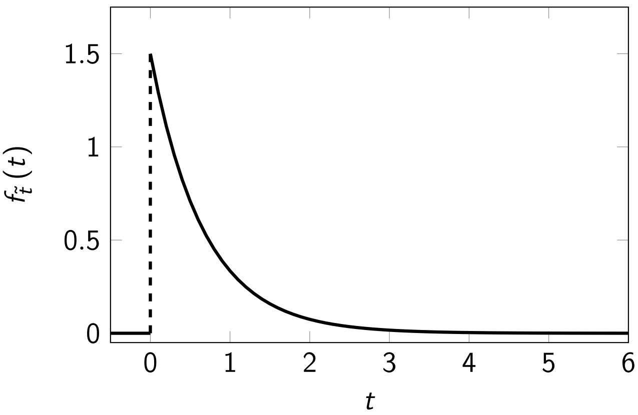
$\lambda = 0.5$



$\lambda = 1$



$\lambda = 1.5$



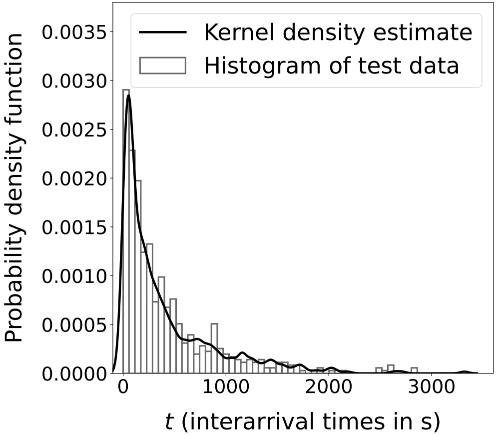
Call center in bank

Goal: Model time between calls (6 am-7 am on weekdays)

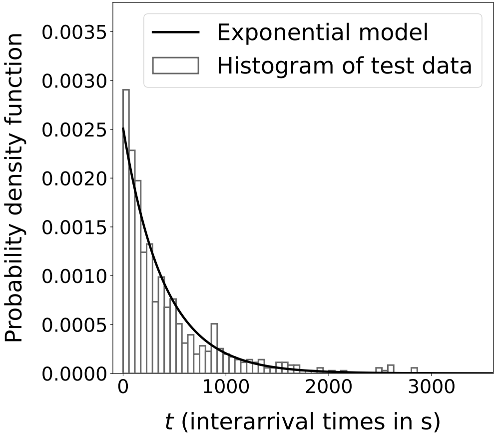
Training set: Calls from January-June 1999

Test set: Calls from July-December 1999

KDE estimate



Exponential model

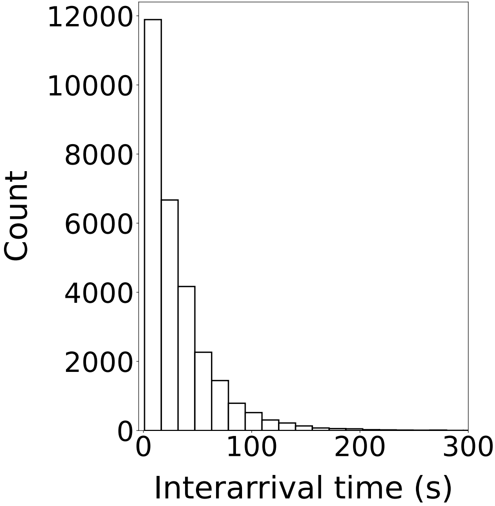


Conditional probabilities

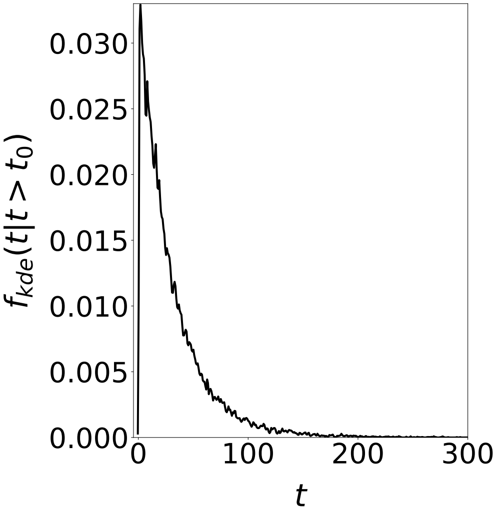
If $\tilde{t} > t_0$ how does the distribution change?

We look at calls between 9 and 10 am on weekdays

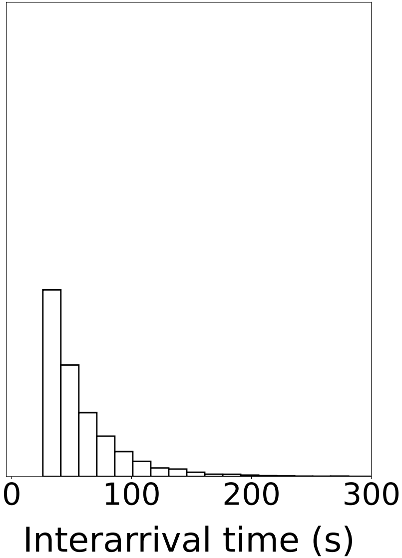
Histogram $t_0 = 0$



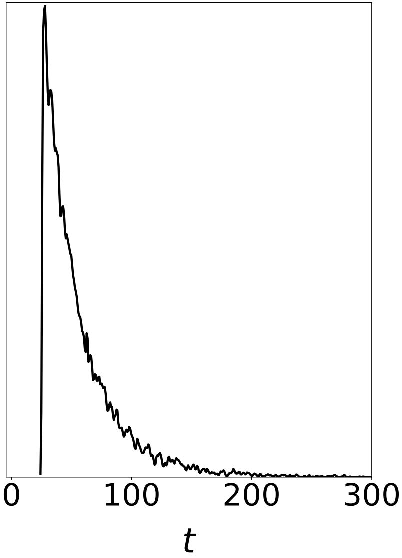
KDE $t_0 = 0$



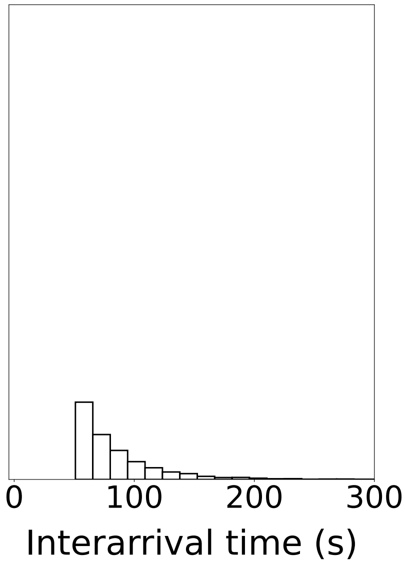
Histogram $t_0 = 25$ s



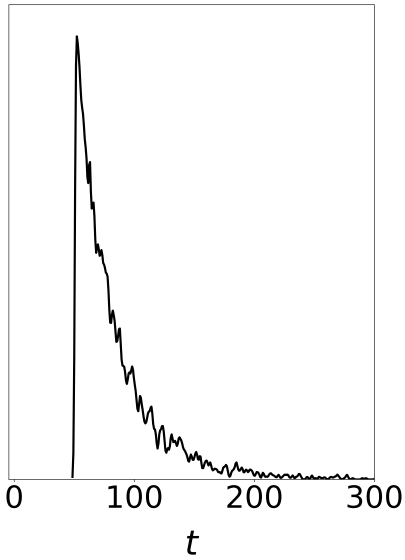
KDE $t_0 = 25$ s



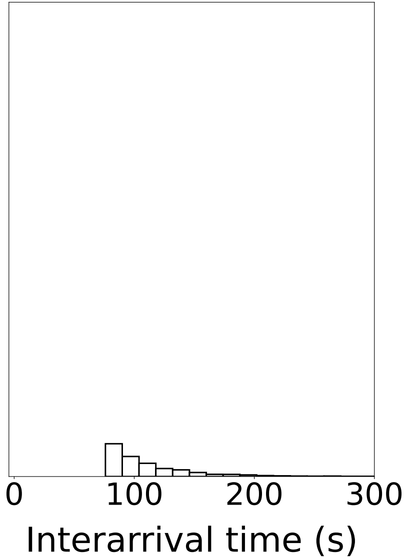
Histogram $t_0 = 50$ s



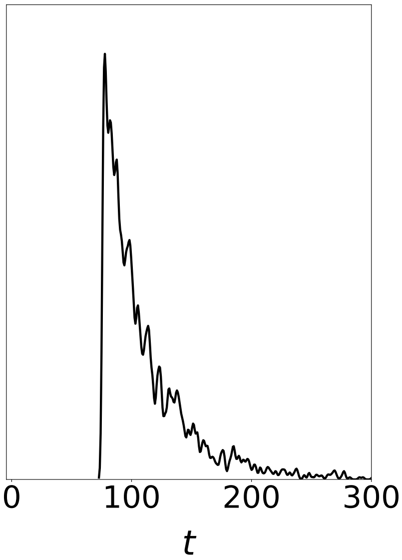
KDE $t_0 = 50$ s



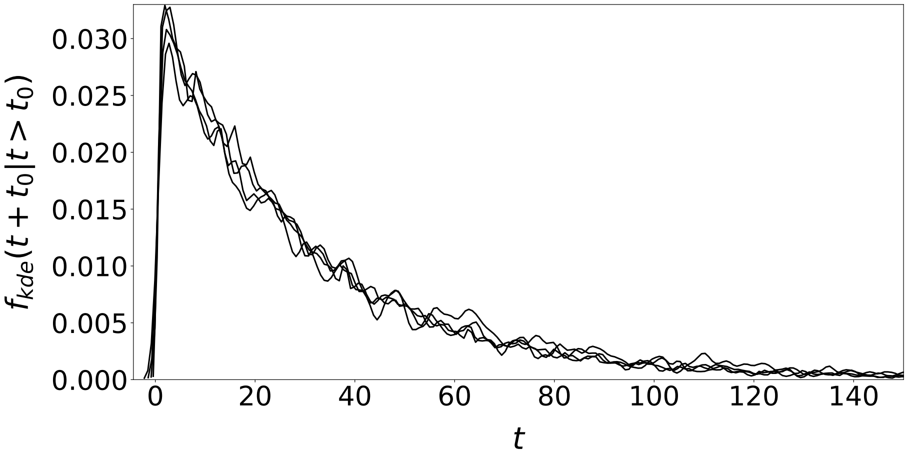
Histogram $t_0 = 75$ s



KDE $t_0 = 75$ s



Densities are similar

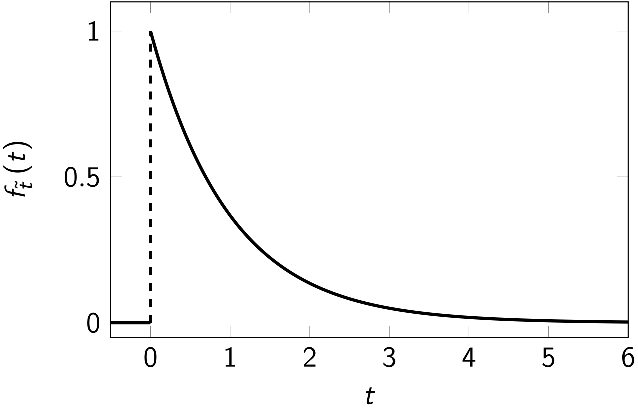


Conditional pdf given $\tilde{t} > t_0$

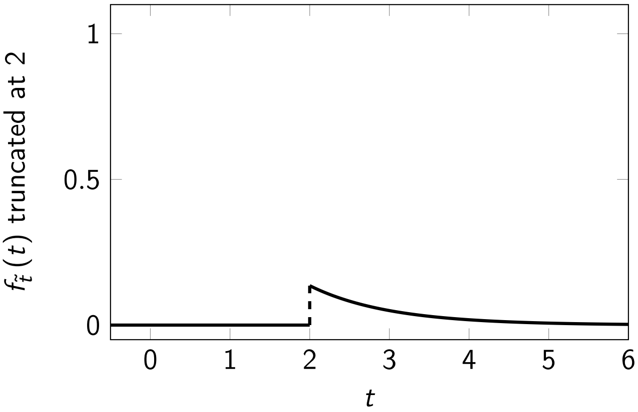
$$
\begin{aligned}
F_{\tilde{t}\,|\,\tilde{t}>t_0}(t) &= \mathrm{P}\left(\tilde{t} \leq t \,|\, \tilde{t} > t_0\right) \\
&= \frac{\mathrm{P}\left(t_0 < \tilde{t} \leq t\right)}{\mathrm{P}\left(\tilde{t} > t_0\right)} \\
&= \frac{F_{\tilde{t}}(t) - F_{\tilde{t}(t_0)}}{1 - F_{\tilde{t}}(t_0)} \\
&= \frac{e^{-\lambda t_0} - e^{-\lambda t}}{e^{-\lambda t_0}} \\
&= 1 - e^{-\lambda(t-t_0)} \\
f_{\tilde{t}|\tilde{t}>t_0}(t) &= \lambda e^{-\lambda(t-t_0)}
\end{aligned}
$$

Exponential starting at t_0! Exponential distribution is memoryless

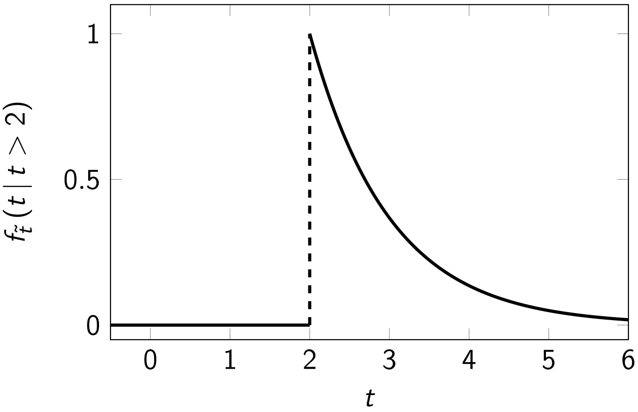
Graphical explanation



Graphical explanation



Graphical explanation



What have we learned?

Derivation of the exponential distribution

The exponential distribution is memoryless