Lecture 10 – Solar Prediction

Dr. Yu Zhang

ECE, UCSC

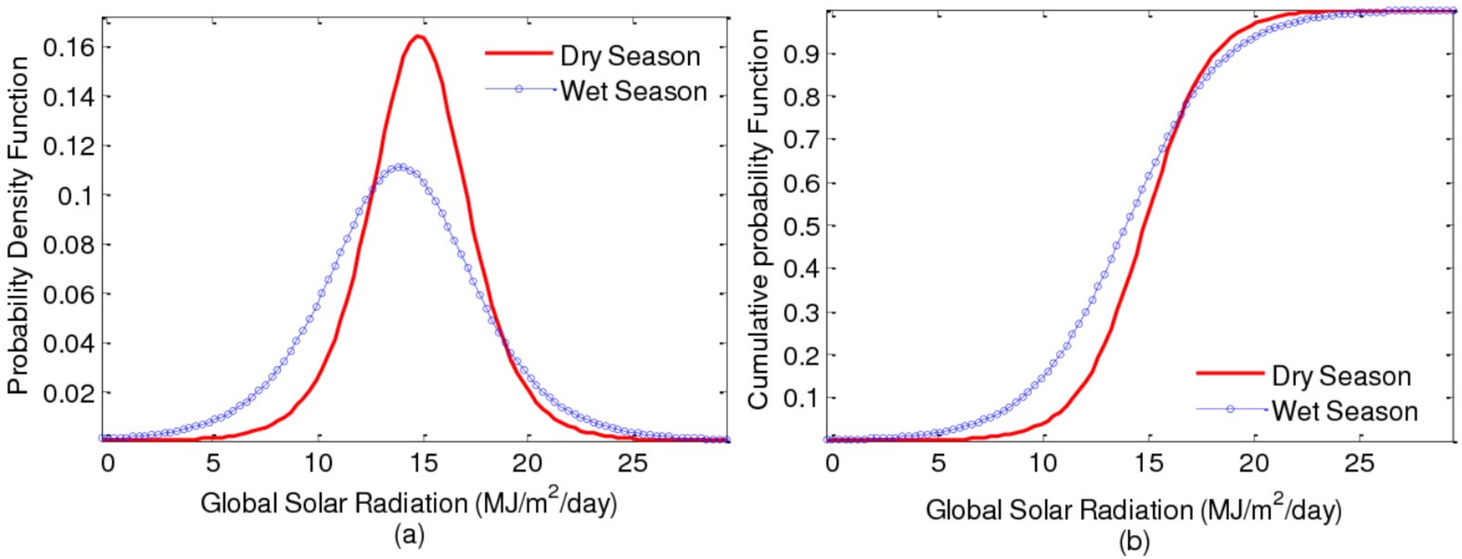
ECE180J

Outline

- Maximum Likelihood Estimation
- Kernel Density Estimation
- PDFs for Solar Irradiation
- Time Series Analysis
- Supervised Learning: Regression & Classification

- Wikipedia

Probability Functions for Solar Irradiation



Depending on location & time, commonly-used probability functions of hourly solar irradiation are listed as follows

- Exponential distribution
- Normal distribution
- Lognormal distribution
- Weibull distribution
- Gamma distribution
- Beta distribution

Maximum Likelihood Estimation (MLE)

MLE: Given n data samples, estimate the parameters of a probability distribution by maximizing a likelihood function, so that the observed data are most probable.

The likelihood measures the goodness of fit for a model. It is formed from the joint PDF of the samples:

1) treat the random variables as fixed at the observed values
2) viewed and used as a function of the parameters only.

It is often convenient the maximize the log-likelihood:

$$\ell(\theta\,;\mathbf{y}) = \ln L_n(\theta\,;\mathbf{y})$$

A necessary 1st order optimality condition of MLE:

$$\frac{\partial \ell}{\partial \theta_1} = 0, \quad \frac{\partial \ell}{\partial \theta_2} = 0, \quad \ldots, \quad \frac{\partial \ell}{\partial \theta_k} = 0$$

Kernel Density Estimation (KDE)

KDE: A non-parametric way to estimate the pdf of a random variable based on kernels as weights.

1. Let $(x_1, x_2, \ldots, x_n)$ be i.i.d. samples drawn from some univariate distribution with an unknown density f at any given point x.
2. We are interested in estimating the shape of this function f.
3. Its *kernel density estimator* is given as

$$\widehat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big)$$

where K is the kernel — a non-negative function
$h > 0$ is a smoothing parameter called *bandwidth*

- Commonly-used kernel functions: uniform, triangular, biweight, triweight, normal, and others.

KDE (Cont'd)

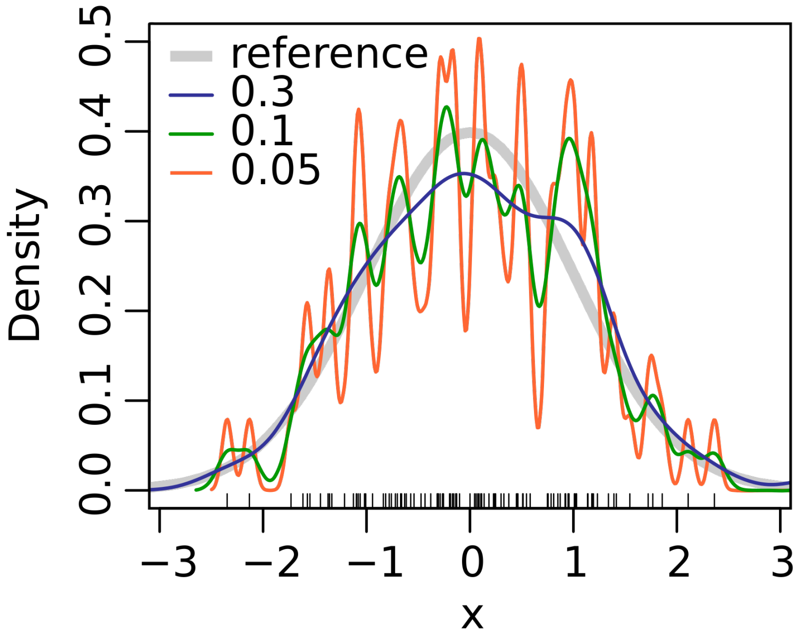


Fig (a): Kernel density estimation of 100 normally distributed random numbers using different smoothing bandwidths.

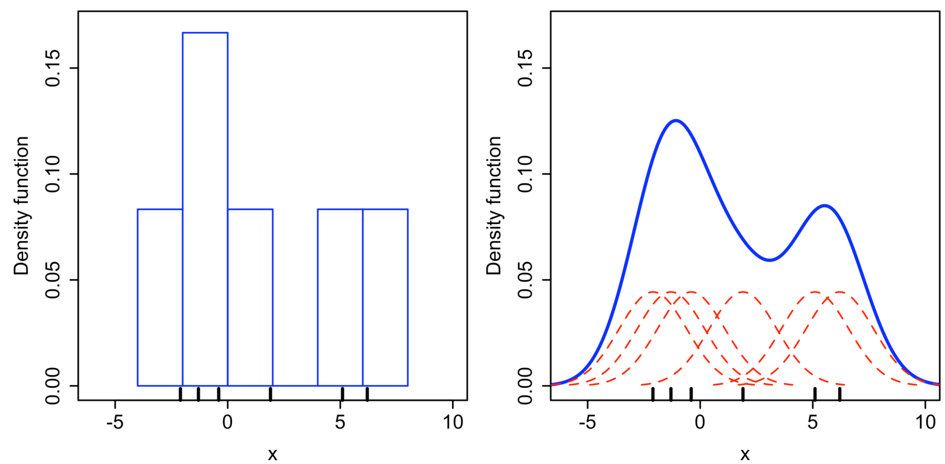
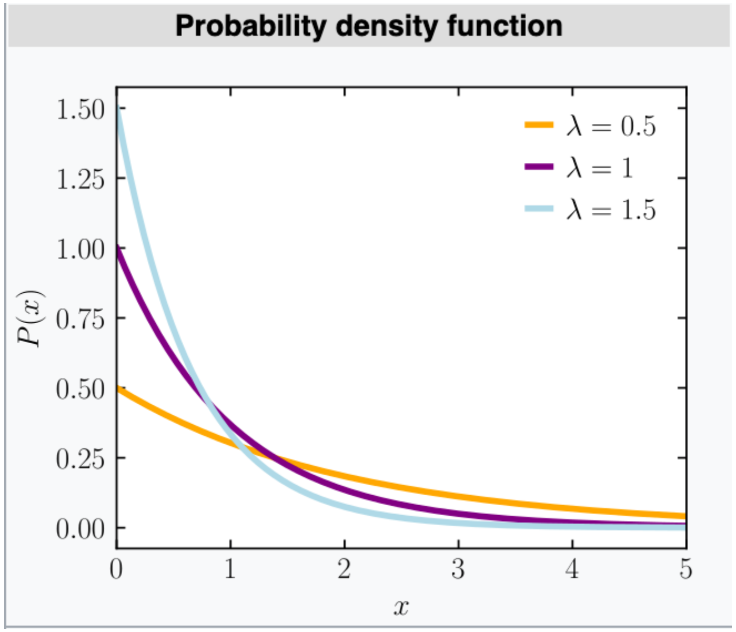
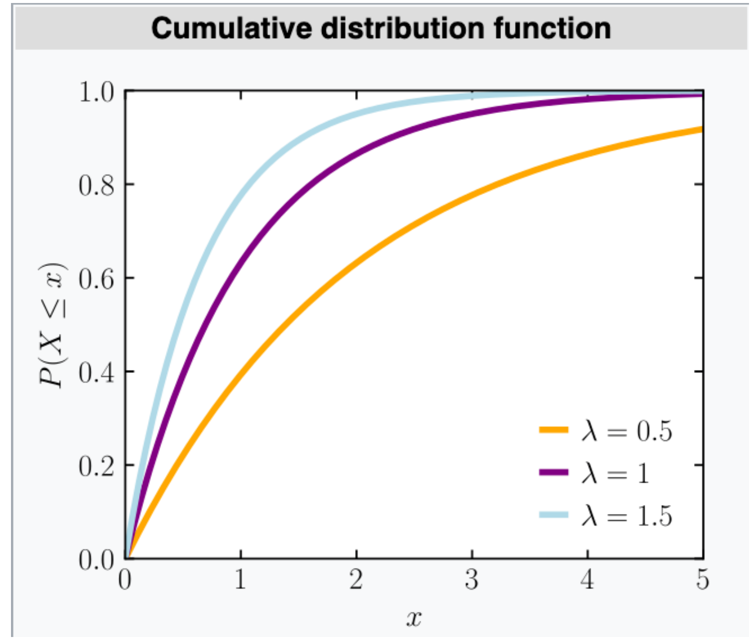


Fig (b): Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The six individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot on the horizontal axis.

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





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

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Expectation, variance, and moments:

$$\mathrm{E}[X] = \frac{1}{\lambda} \qquad \mathrm{Var}[X] = \frac{1}{\lambda^2} \qquad \mathrm{E}[X^n] = \frac{n!}{\lambda^n}$$

Memoryless property:

$$\Pr\left(T > s + t \mid T > s\right) = \Pr(T > t), \qquad \forall s, t \geq 0.$$

MLE for Exponential Distribution

- Given independent and identically distributed (i.i.d.) samples x = (x_1, ..., x_n) drawn from the exp distribution, the likelihood function is given as

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right) = \lambda^n \exp(-\lambda n \bar{x})$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the sample mean.

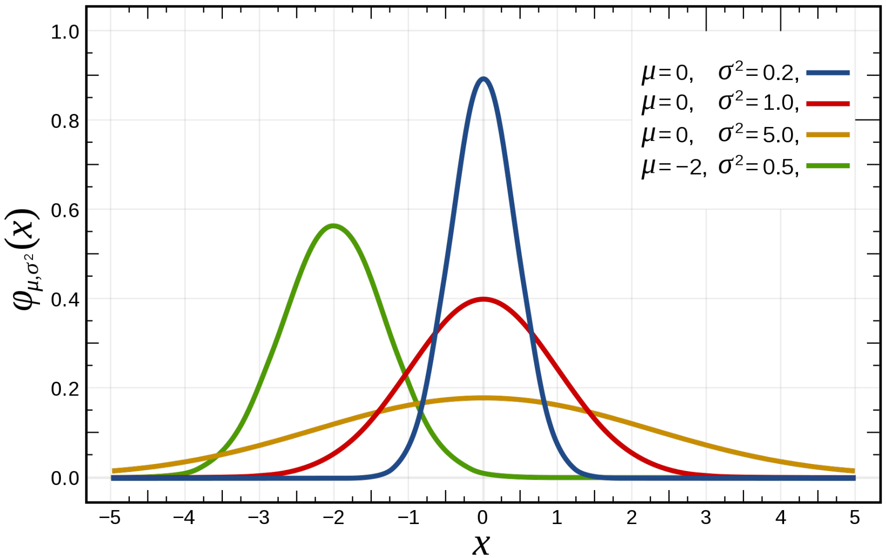
- The derivative of the log-likelihood is:

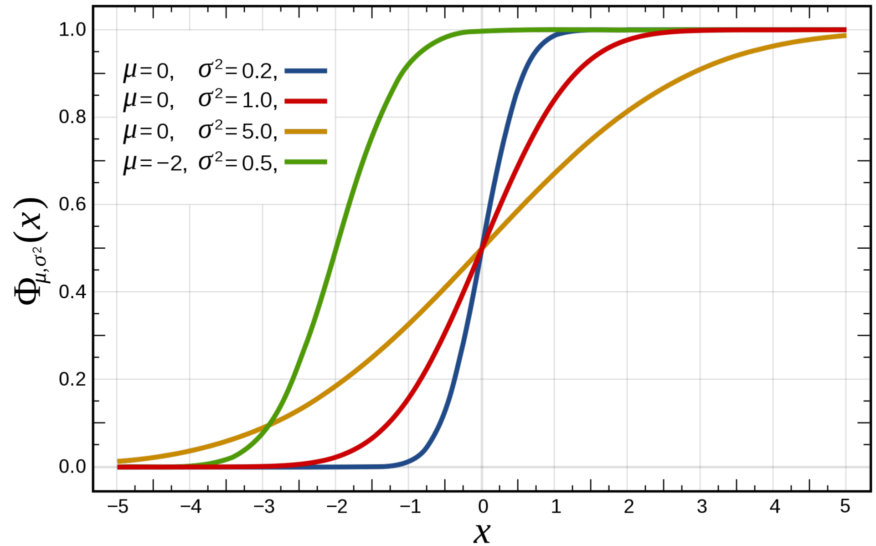
$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{d}{d\lambda} (n \ln \lambda - \lambda n \bar{x}) = \frac{n}{\lambda} - n\bar{x} \begin{cases} > 0, & 0 < \lambda < \frac{1}{\bar{x}}, \\ = 0, & \lambda = \frac{1}{\bar{x}}, \\ < 0, & \lambda > \frac{1}{\bar{x}}. \end{cases}$$

- Consequently, the MLE for the rate parameter is:

$$\widehat{\lambda}_{\text{mle}} = \frac{1}{\bar{x}} = \frac{n}{\sum_i x_i}$$

Normal Distribution $X \sim \mathcal{N}(\mu, \sigma^2)$





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

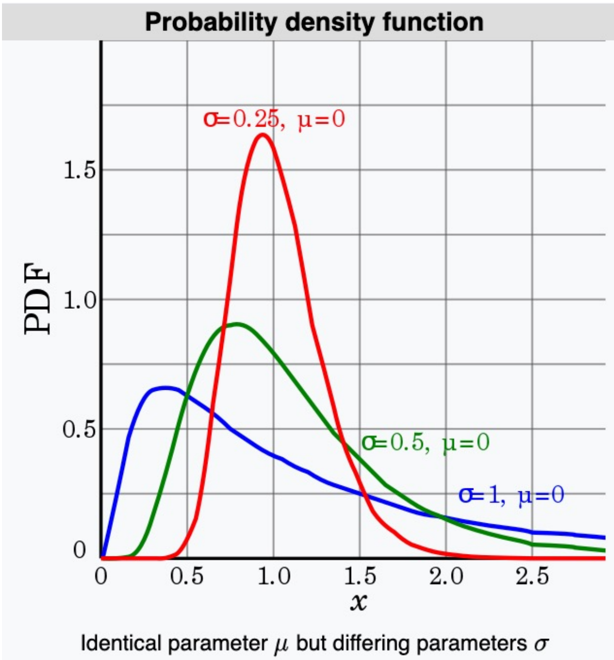
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

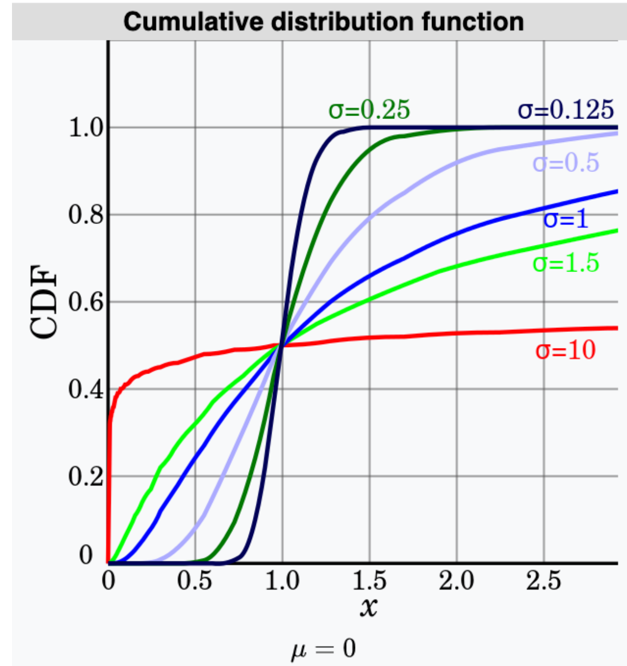
Central moments: $$\mathrm{E}[(X-\mu)^p] = \begin{cases} 0 & \text{if } p \text{ is odd,} \\ \sigma^p (p-1)!! & \text{if } p \text{ is even.} \end{cases}$$

Log-likelihood $$\ln \mathcal{L}(\mu, \sigma^2) = \sum_{i=1}^{n} \ln f(x_i \mid \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

MLE: $$\hat{\mu} = \overline{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Lognormal Distribution $X \sim \text{Lognormal}(\mu_x, \sigma_x^2)$ $\ln(X) \sim \mathcal{N}(\mu, \sigma^2)$



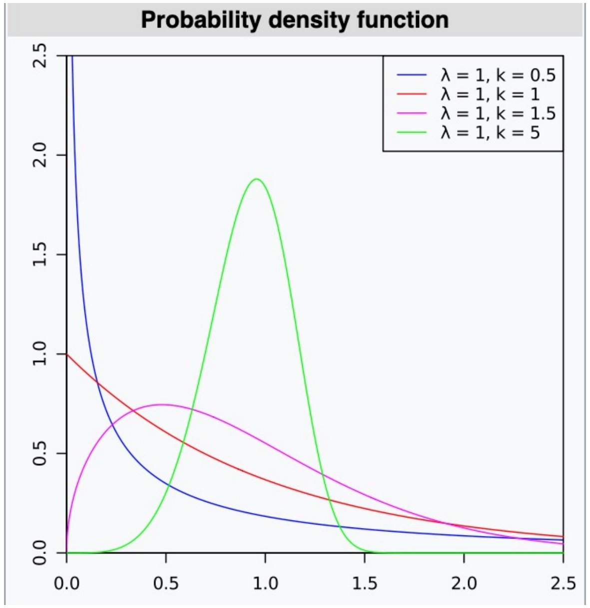


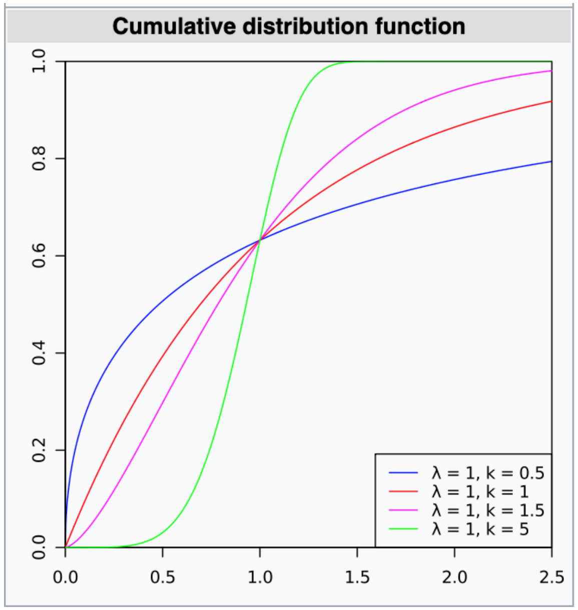
$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right).$$

$$F_X(x) = \Phi\left(\frac{(\ln x) - \mu}{\sigma}\right)$$

MLE: $$\widehat{\mu} = \frac{\sum_k \ln x_k}{n}, \qquad \widehat{\sigma}^2 = \frac{\sum_k (\ln x_k - \widehat{\mu})^2}{n}$$

Weibull Distribution $X \sim \mathrm{Weibull}(\lambda, k)$





$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

where $k > 0$ is the *shape parameter* and $\lambda > 0$ is the *scale parameter*

$$F(x;k,\lambda) = 1 - e^{-(x/\lambda)^k}$$

for $x \geq 0$, and $F(x; k; \lambda) = 0$ for $x < 0$.

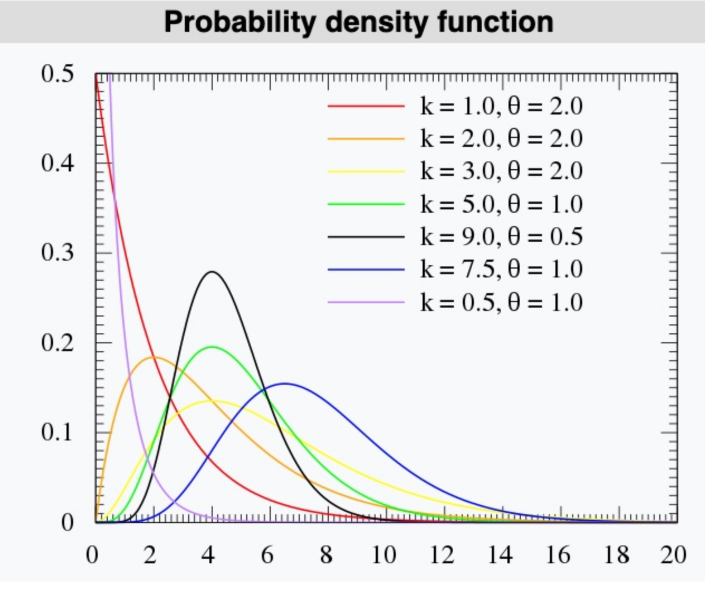
MLE for lambda given k:

$$\widehat{\lambda}^k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

MLE for k is the solution for k of the following equation:

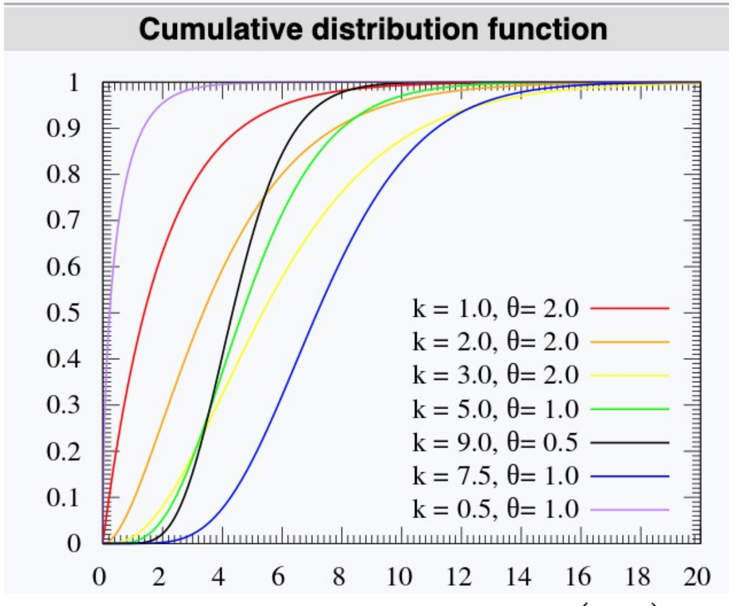
$$0 = \frac{\sum_{i=1}^n x_i^k \ln x_i}{\sum_{i=1}^n x_i^k} - \frac{1}{k} - \frac{1}{n}\sum_{i=1}^n \ln x_i$$

Gamma Distribution $X \sim \Gamma(k, \theta) \equiv \mathrm{Gamma}(k, \theta)$



$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}$$

for $x > 0$ and $k, \theta > 0$.



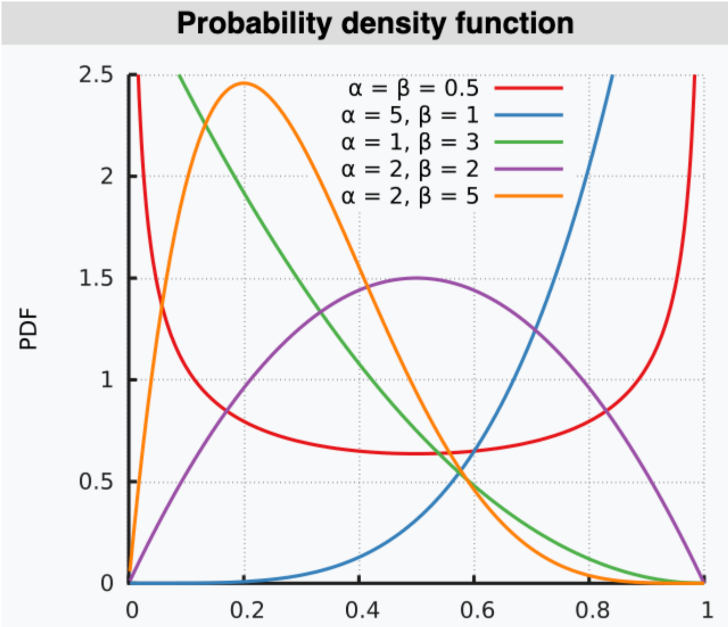
$$F(x; k, \theta) = \int_0^x f(u; k, \theta)\, du = \frac{\gamma\left(k, \frac{x}{\theta}\right)}{\Gamma(k)},$$

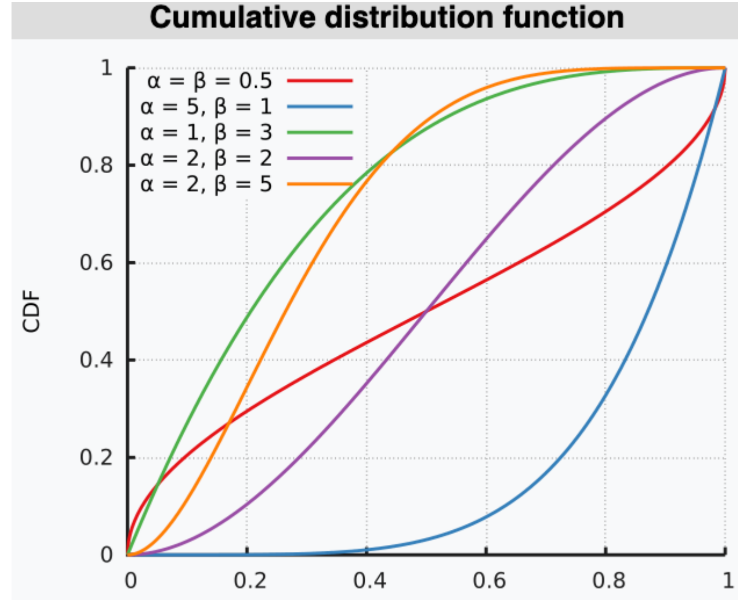
where $\gamma\left(k, \frac{x}{\theta}\right)$ is the lower incomplete gamma function.

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx, \qquad \Re(z) > 0\,. \qquad \Gamma(n) = (n-1)!$$

Log-likelihood: $\ell(k, \theta) = (k-1) \sum_{i=1}^{N} \ln(x_i) - \sum_{i=1}^{N} \frac{x_i}{\theta} - Nk\ln(\theta) - N\ln(\Gamma(k))$

Beta Distribution $X \sim \mathrm{Beta}(\alpha, \beta)$





$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$$F(x; \alpha, \beta) = \frac{\mathrm{B}(x; \alpha, \beta)}{\mathrm{B}(\alpha, \beta)} = I_x(\alpha, \beta)$$

where $\mathrm{B}(x; \alpha, \beta)$ is the incomplete beta function

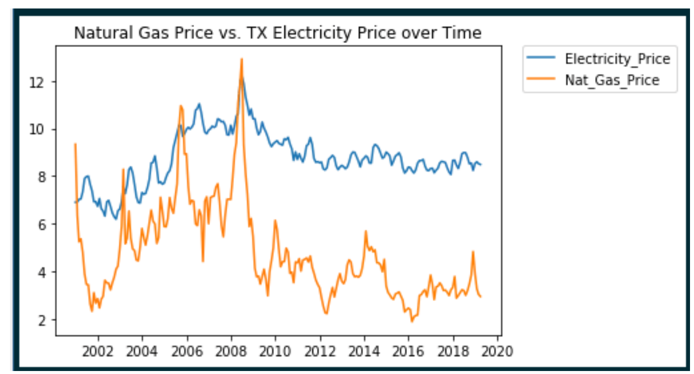
$I_x(\alpha, \beta)$ is the regularized incomplete beta function.

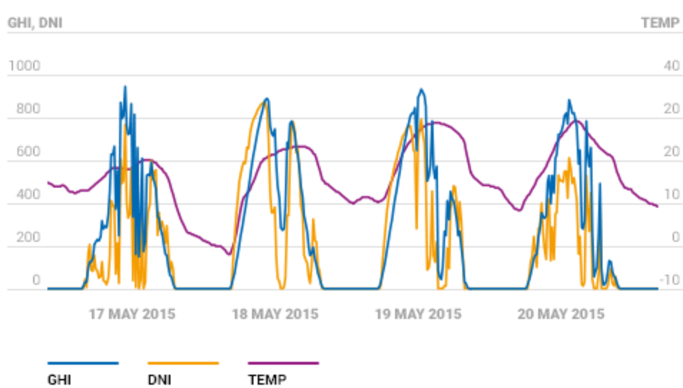
Log-likelihood function:

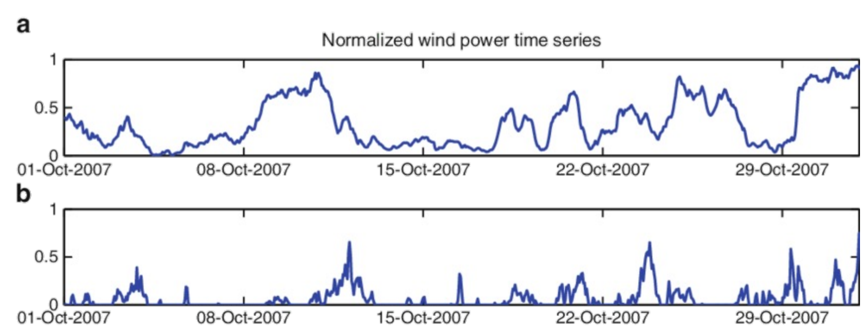
$$\ln \mathcal{L}(\alpha, \beta \mid X) = (\alpha - 1)\sum_{i=1}^{N} \ln(X_i) + (\beta - 1)\sum_{i=1}^{N} \ln(1 - X_i) - N \ln \mathrm{B}(\alpha, \beta)$$

Time Series

- A time series is a series of data points indexed in time order.
- A sequence of discrete-time data: often taken at successive equally spaced points in time.
- Examples: heights of ocean tides, counts of sunspots, stock prices, solar/wind power generation, etc.







ARMA

Auto Regressive Moving Average (ARMA): A class of models that ‘explains’ a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

Given time series data X_t where t is an integer index and the X_t are real numbers, an $\mathrm{ARMA}(p', q)$ model is given by

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_{p'} X_{t-p'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

L is the lag operator,

α_i are the parameters of the autoregressive part of the model,

θ_i are the parameters of the moving average part

The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

Auto Regressive Integrated Moving Average (ARIMA):

- A stationary time series' properties do not depend on the time at which the series is observed.
- For a wide-sense stationary time series, the mean and the variance/autocovariance keep constant over time.
- Differencing: A transformation applied to a non-stationary time-series to make it stationary in the mean sense (i.e., to remove the non-constant trend).

1st-order differencing: $y'_t = y_t - y_{t-1}$

2nd-order differencing:

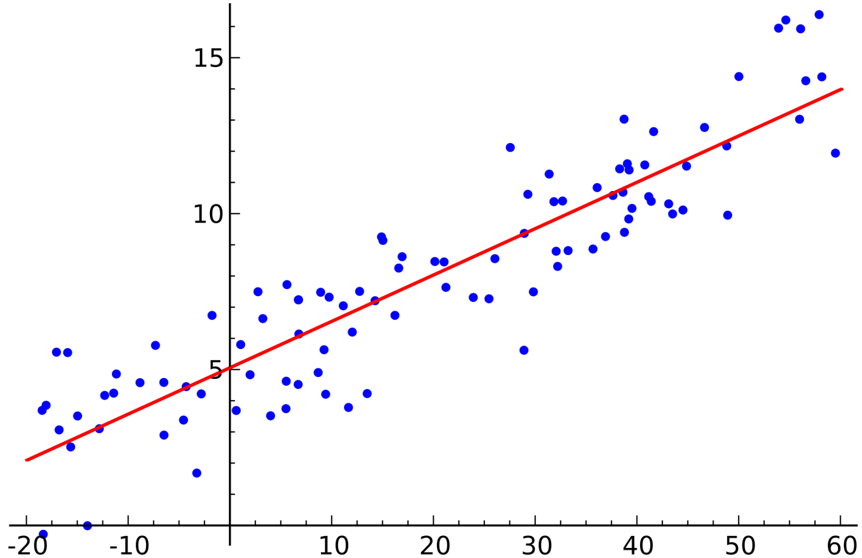
$$\begin{aligned} y^*_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \end{aligned}$$

An ARIMA(p,d,q) process is given by

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

Regression Analysis

- Regression analysis: A set of statistical processes for estimating the relationships between dependent variables (‘outcome variables’) and independent variables (aka, 'predictors', 'covariates', or 'features’).
- Linear regression: Find the line that most closely fits the data according to a specific mathematical criterion.
- For example, the method of ordinary least squares computes the unique line/hyperplane that minimizes the sum of squared differences between the true data and that line/ hyperplane.



Regression Model

General regression model:

$$Y_i = f(X_i, \beta) + e_i$$

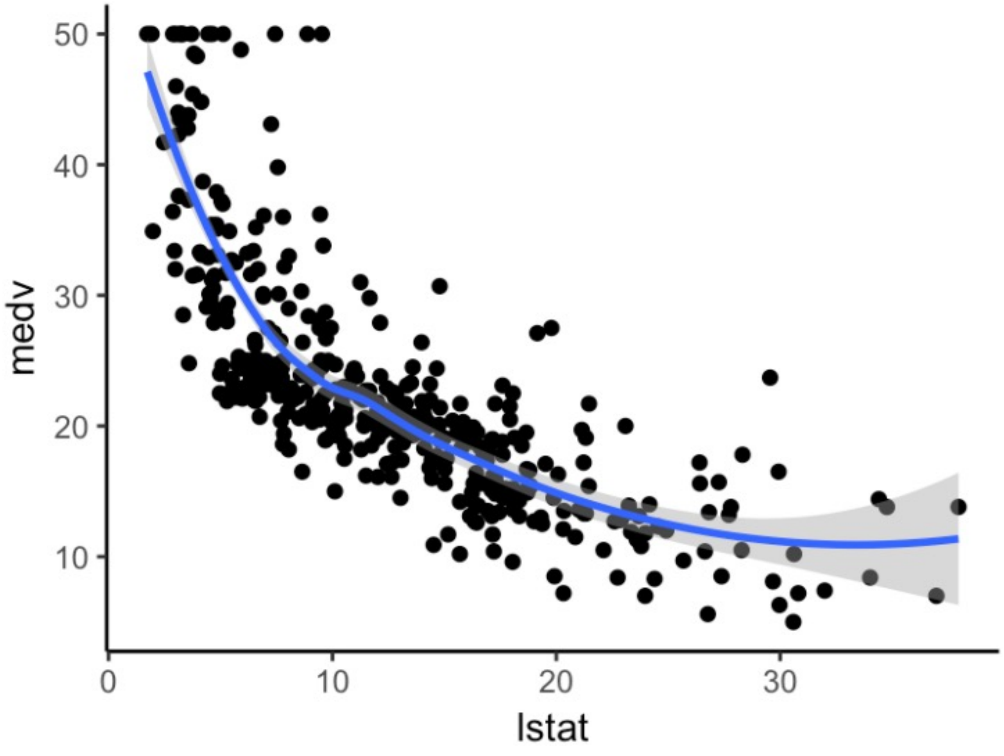
- The **unknown parameters**, often denoted as a scalar or vector β.
- The **independent variables**, which are observed in data and are often denoted as a vector X_i
- The **dependent variable**, which are observed in data and often denoted using the scalar Y_i.
- The **error terms**, which are *not* directly observed in data and are often denoted using the scalar e_i.

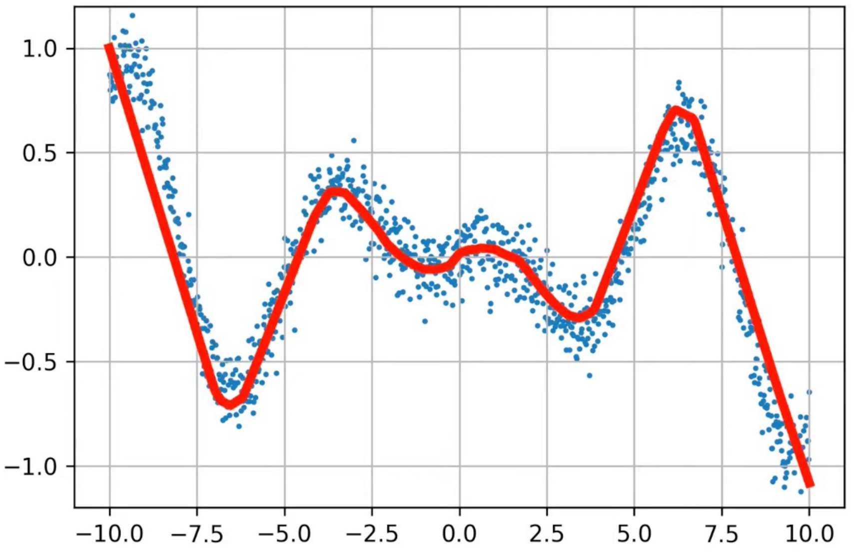
A linear regression model: $Y_i = \beta_0 + \beta_1 X_i + e_i$

β is estimated by solving the least-squares problem:

$$\beta^* = \operatorname{argmin}_\beta \sum_{i=1}^{N} (Y_i - f(X_i, \beta))^2$$

Nonlinear Regression

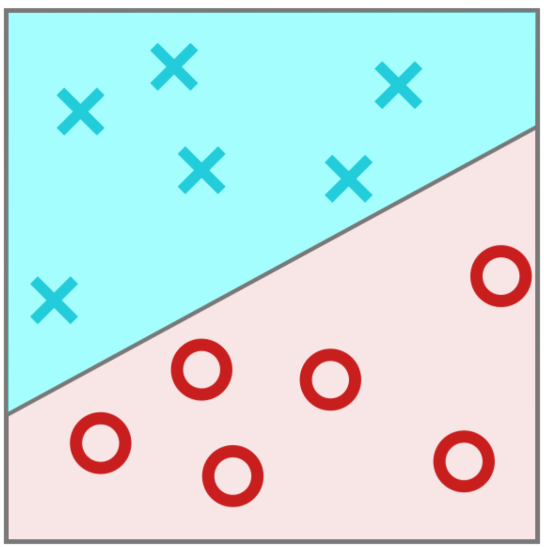




Supervised Learning

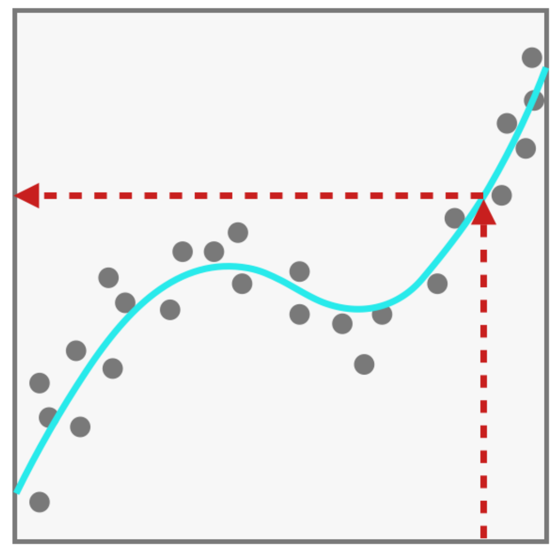
Supervised learning (Classification and Regression) is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.

Classification Groups observations into "classes"



Here, the line classifies the observations into X's and O's

Regression predicts a numeric value



Here, the fitted line provides a predicted output, if we give it an input