Centro Brasileiro de Pesquisas Físicas

Métodos para Análise de grande volume de dados e Astroinformática

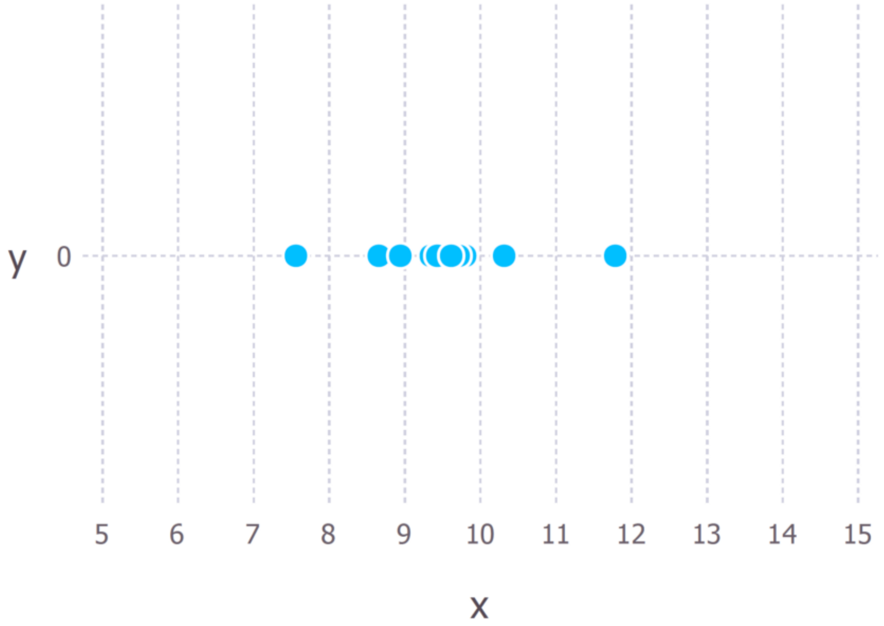
Clécio Roque De Bom – debom@cbpf.br

clearnightsrthebest.com

CBPF

Maximum Likelihood Estimation

Maximum likelihood estimation is a method that determines values for the parameters of a GIVEN model.



<u>Independent and identically distributed random variables</u>

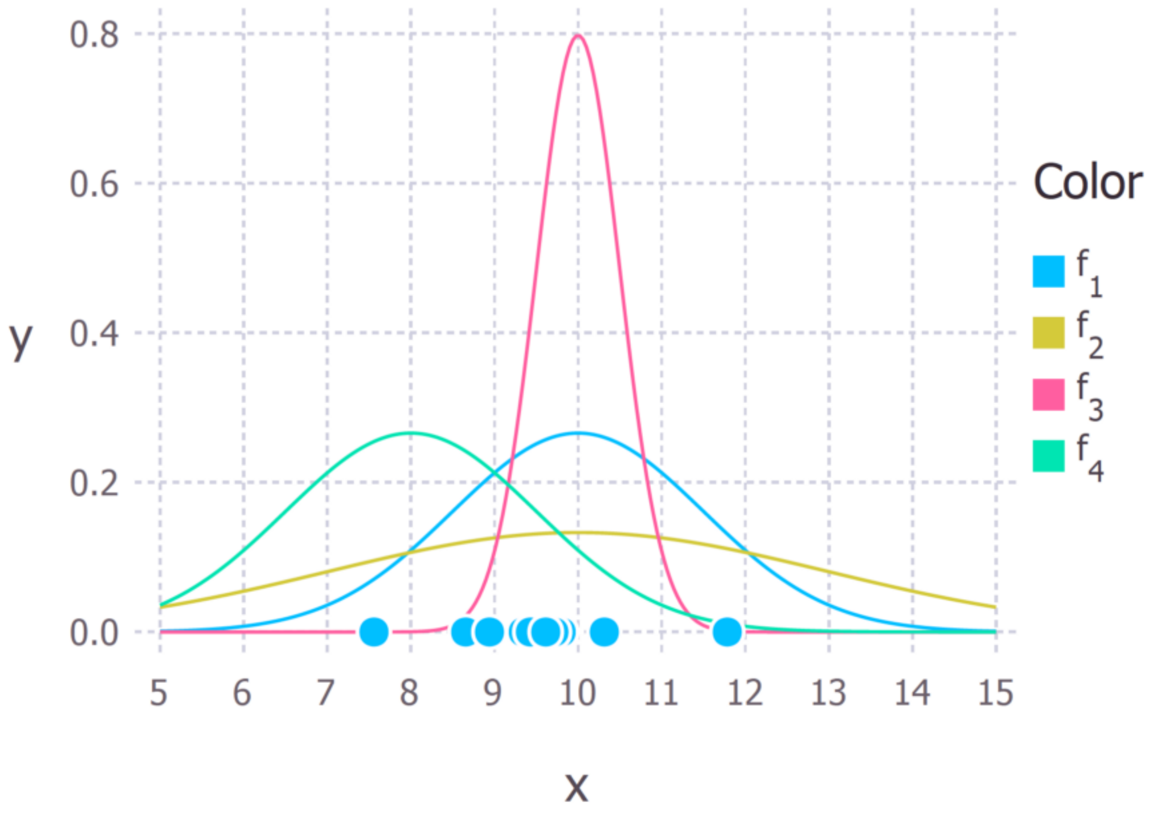
the i.i.d. assumption requires that the observation of any given data point does not depend on the observation of any other data point (each gathered data point is an independent experiment) and that each data point is generated from same distribution family with the same parameters.

Figure credit, adapted from: https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1

Maximum Likelihood Estimation

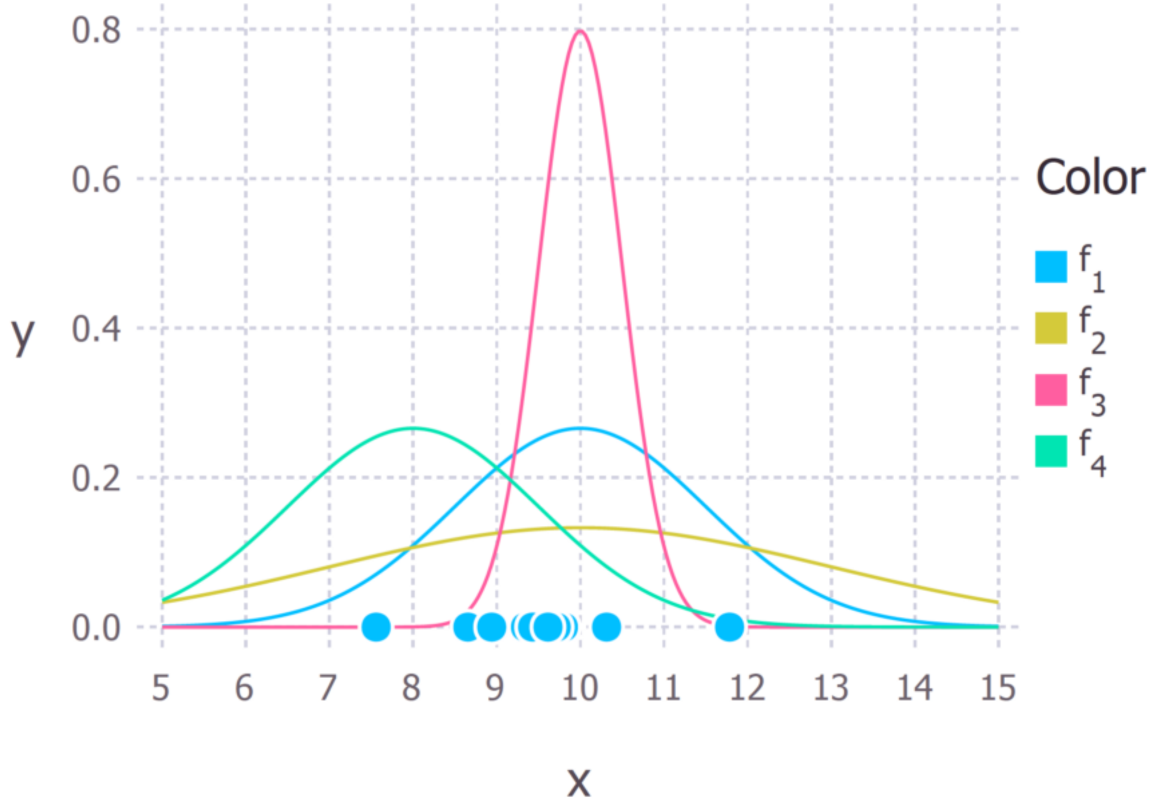
Maximum likelihood estimation is a method that determines values for the parameters of a GIVEN model.

Assume that the data generation process can be described by a (approximate) Gaussian distribution.



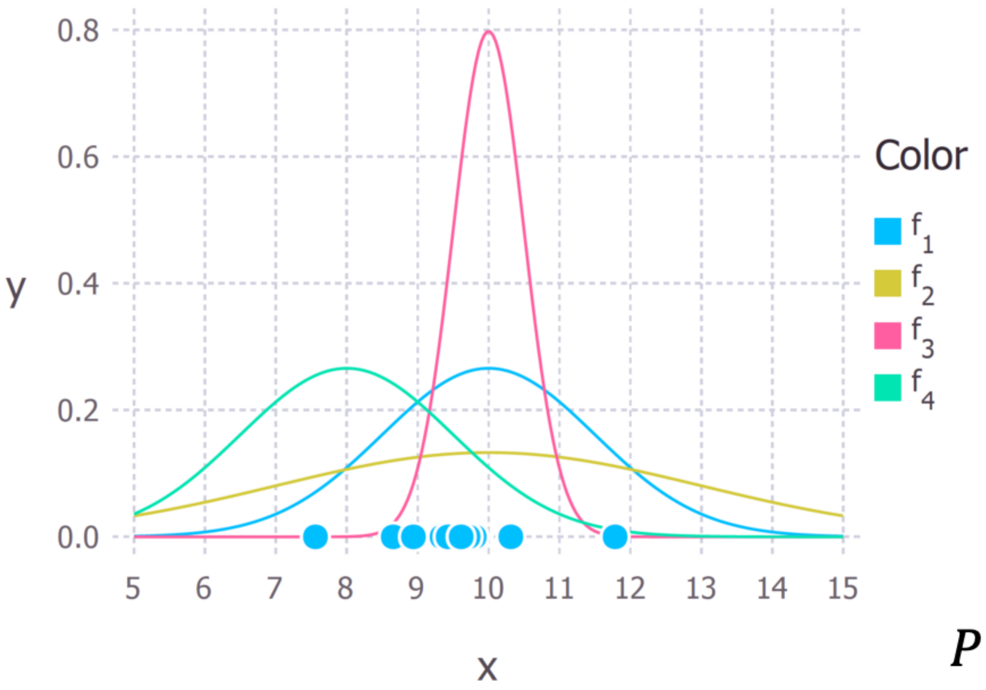
Maximum Likelihood Estimation

Assume that the data generation process can be described by a (approximate) Gaussian distribution.



$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Maximum Likelihood Estimation



$$P(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

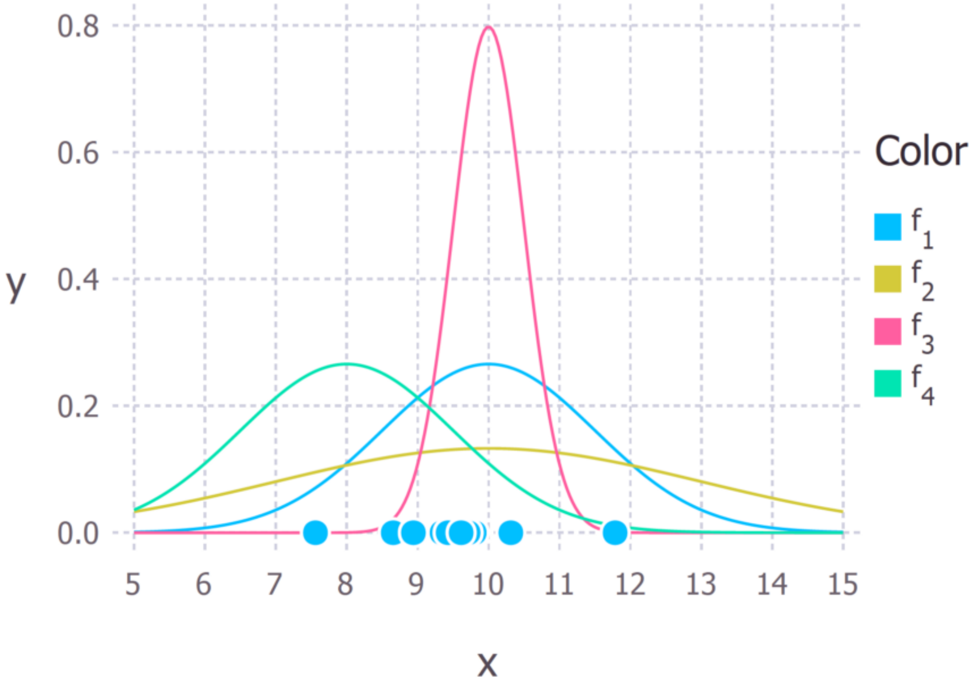
$$P(9,9.5,11;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(9-\mu)^2}{2\sigma^2}\right)\times\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(9.5-\mu)^2}{2\sigma^2}\right)\times\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(11-\mu)^2}{2\sigma^2}\right)$$

$$P(x_1,x_2,\ldots x_n|\theta)\rightarrow f(x_1,x_2,\ldots x_n|\theta)$$

Now, all we need to do is to maximize it for the parameters!

$$\hat{\theta}_{MLE}=argmax_{\theta}\prod_{i}^{n}f(x_i|\theta)$$

Maximum Likelihood Estimation



Now, all we need to do is to maximize it for the parameters!

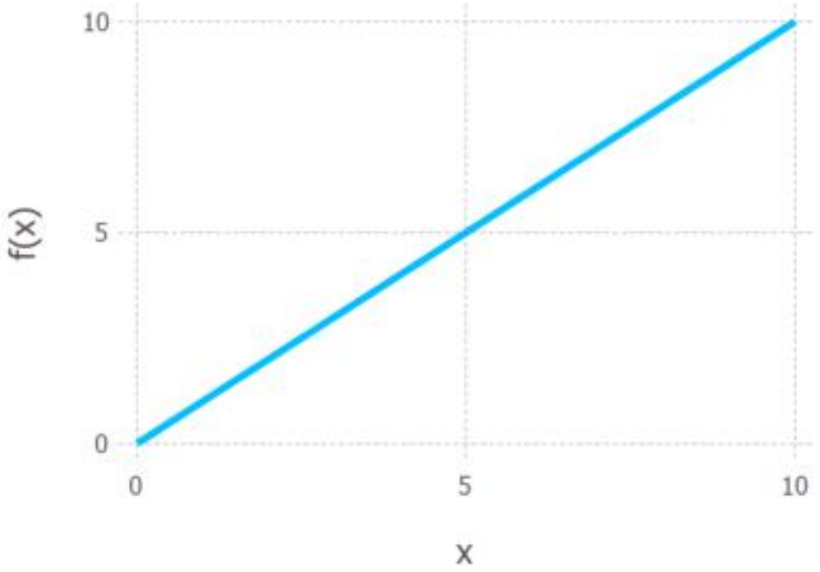
$$\hat{\theta}_{MLE} = argmax_{\theta} \prod_{i}^{n} f(x_i|\theta)$$

$$argmax_{\theta} \prod_{i}^{n} f(x_i|\theta) \rightarrow \frac{\partial}{\partial\theta} \prod_{i}^{n} f(x_i|\theta) = 0$$

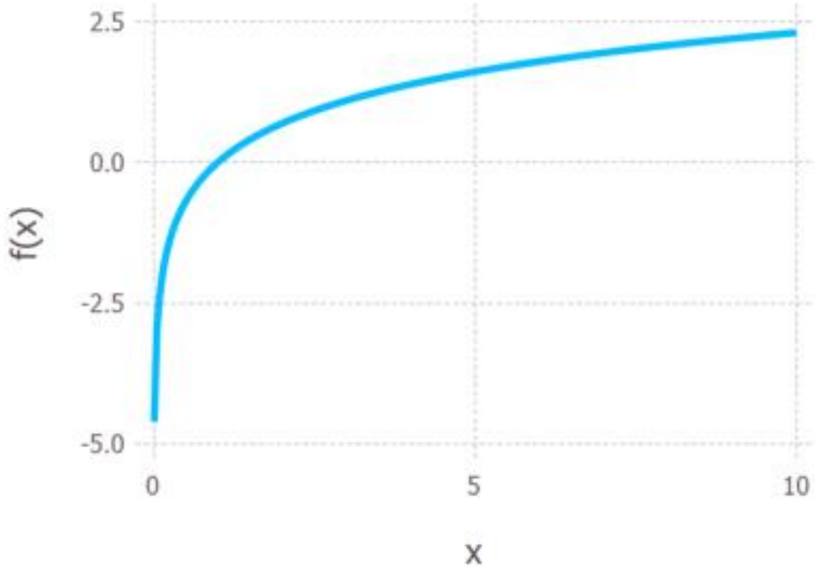
$$\frac{\partial}{\partial\theta} \prod_{i}^{n} f(x_i|\theta) \sim \frac{\partial}{\partial\theta} \ln \langle \prod_{i}^{n} f(x_i|\theta) \rangle = \frac{\partial}{\partial\theta} \sum_{i}^{n} \ln \langle f(x_i|\theta) \rangle$$

$$= \sum_{i}^{n} \frac{\partial}{\partial\theta} \ln \langle f(x_i|\theta) \rangle = 0$$

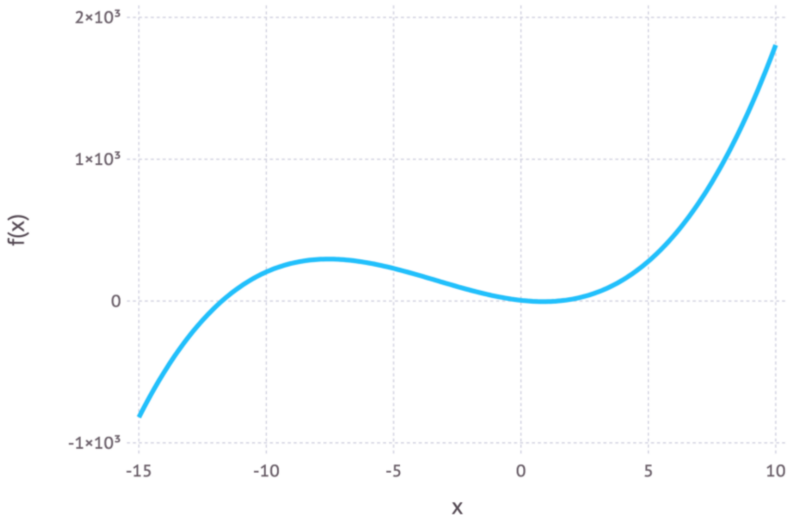
Monotonically inscreasing function



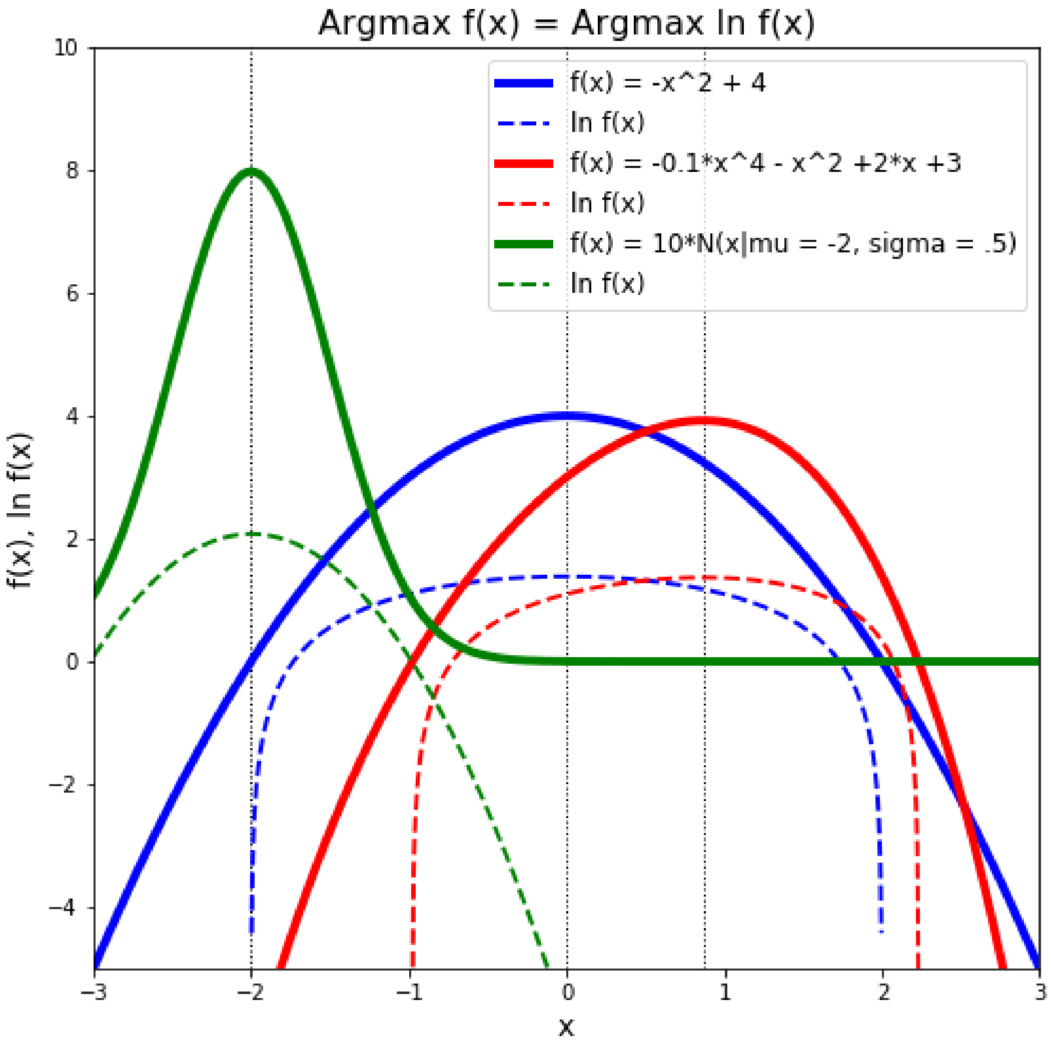
(a) $f(x) = x$



(b) $f(x) = \ln(x)$



Monotonically inscreasing function



MLE Method

$$\sum_{i}^{n} \frac{\partial}{\partial \theta} \ln\langle f(x_i|\theta)\rangle = \sum_{i}^{n} \nabla_{\mu,\sigma} \ln\langle f(x_i|\mu,\sigma)\rangle = 0$$

$$\sum_{i}^{n} \nabla_{\mu} \ln\langle \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} \rangle = \sum_{i}^{n} \nabla_{\mu} -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\sum_{i}^{n} \nabla_{\mu} -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2} = -\frac{1}{2\sigma^2}\sum_{i}^{n} \nabla_{\mu} (x_i-\mu)^2$$

$$= -\frac{1}{2\sigma^2}\sum_{i}^{n} -2(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i}^{n}(x_i-\mu)$$

MLE Method

$$\sum_i^n \frac{\partial}{\partial\theta} \ln\langle f(x_i|\theta)\rangle = \sum_i^n \nabla_{\mu,\sigma} \ln\langle f(x_i|\mu,\sigma)\rangle = 0$$

$$\sum_i^n \nabla_\mu \ln\langle\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}\rangle = \sum_i^n \nabla_\mu - \frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\sum_i^n \nabla_\mu - \frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2} = -\frac{1}{2\sigma^2}\sum_i^n \nabla_\mu (x_i-\mu)^2$$

$$= -\frac{1}{2\sigma^2}\sum_i^n -2(x_i-\mu) = \frac{1}{\sigma^2}\sum_i^n (x_i-\mu)$$

MLE Method

$$\sum_i^n \nabla_\mu \ \ln\langle\frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}\rangle = \sum_i^n \nabla_\mu \ -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i - \ \mu)^2}{2\sigma^2}$$

$$\sum_i^n \nabla_\mu \ -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i - \ \mu)^2}{2\sigma^2} = -\frac{1}{2\sigma^2}\sum_i^n \nabla_\mu \ (x_i - \ \mu)^2$$

$$= \ -\frac{1}{2\sigma^2}\sum_i^n -2(x_i - \ \mu) = \frac{1}{\sigma^2}\sum_i^n (x_i - \ \mu)$$

$$\frac{1}{\sigma^2}\sum_i^n (x_i - \ \mu) = 0 \rightarrow \boxed{\hat{\mu}_{MLE} = \frac{1}{n}\sum_i^n x_i}$$

MLE Method

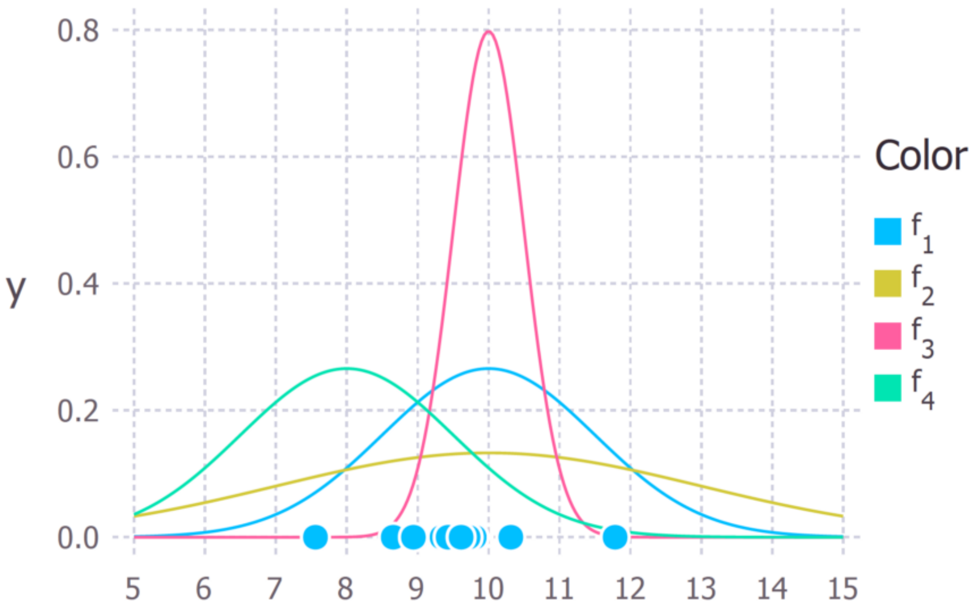
$$\sum_{i}^{n} \nabla_\sigma \ln \langle \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i - \mu)^2}{2\sigma^2}} \rangle = \sum_{i}^{n} \nabla_\sigma - \frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$\sum_{i}^{n} \nabla_\sigma - \frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$= -\frac{n}{2}\nabla_\sigma \ln \sigma^2 - \frac{1}{2}\nabla_\sigma \langle \frac{1}{\sigma^2}\sum_{i}^{n}(x_i - \mu)^2 \rangle$$

$$= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i}^{n}(x_i - \mu)^2$$

Maximum Likelihood Estimation - Example



$$P(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

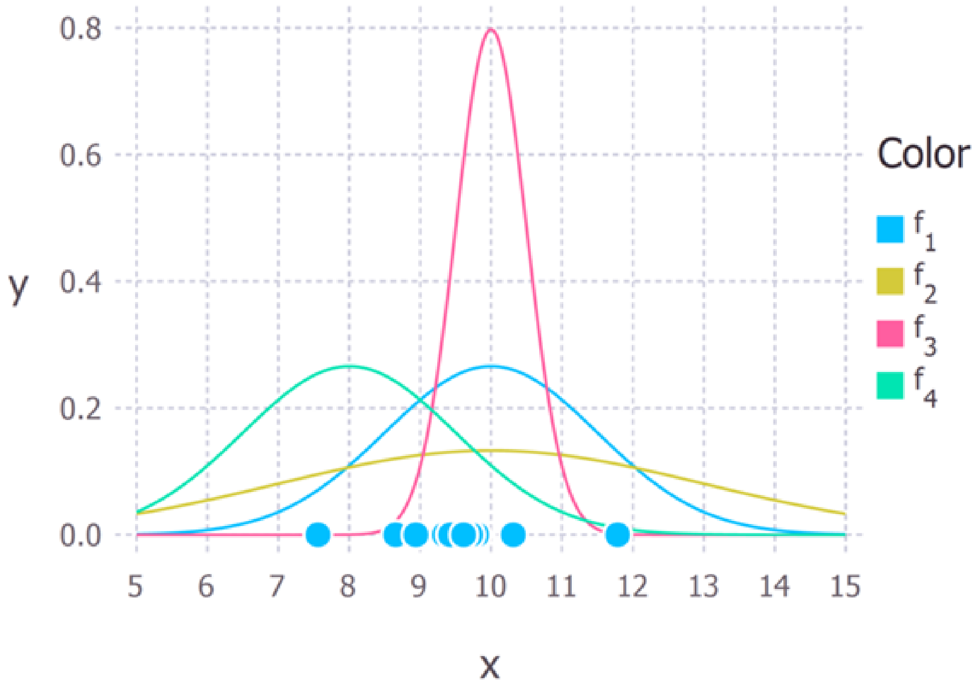
$$P(9, 9.5, 11; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(9-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(9.5-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(11-\mu)^2}{2\sigma^2}\right)$$

$$\ln(P(x;\mu,\sigma)) = \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(9-\mu)^2}{2\sigma^2} + \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(9.5-\mu)^2}{2\sigma^2} + \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(11-\mu)^2}{2\sigma^2}$$

Back to our example with numbers

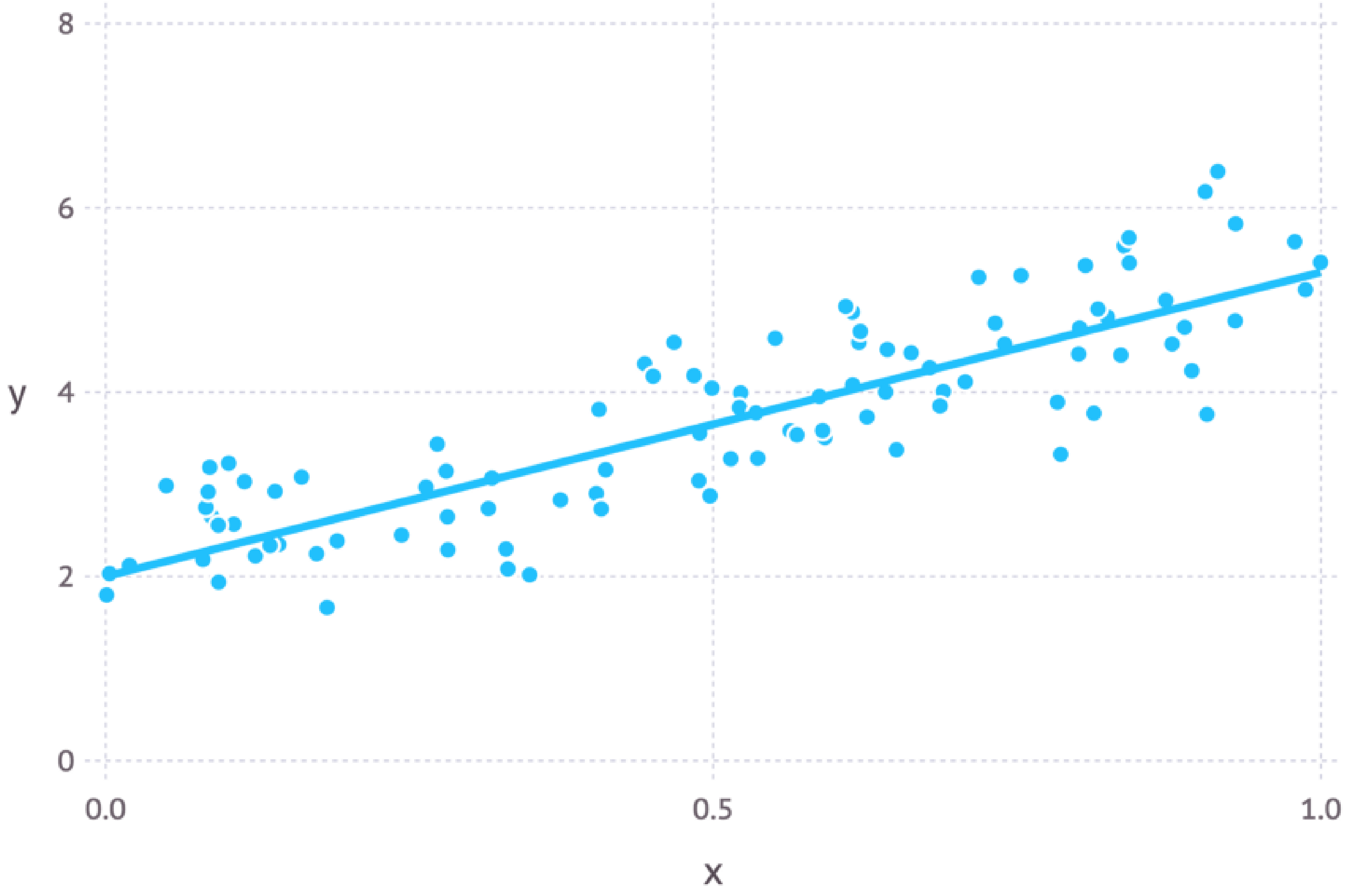
$$\ln(P(x;\mu,\sigma)) = -3\ln(\sigma) - \frac{3}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\left[(9-\mu)^2 + (9.5-\mu)^2 + (11-\mu)^2\right]$$

$$\frac{\partial \ln(P(x;\mu,\sigma))}{\partial \mu} = \frac{1}{\sigma^2}\left[9 + 9.5 + 11 - 3\mu\right].$$

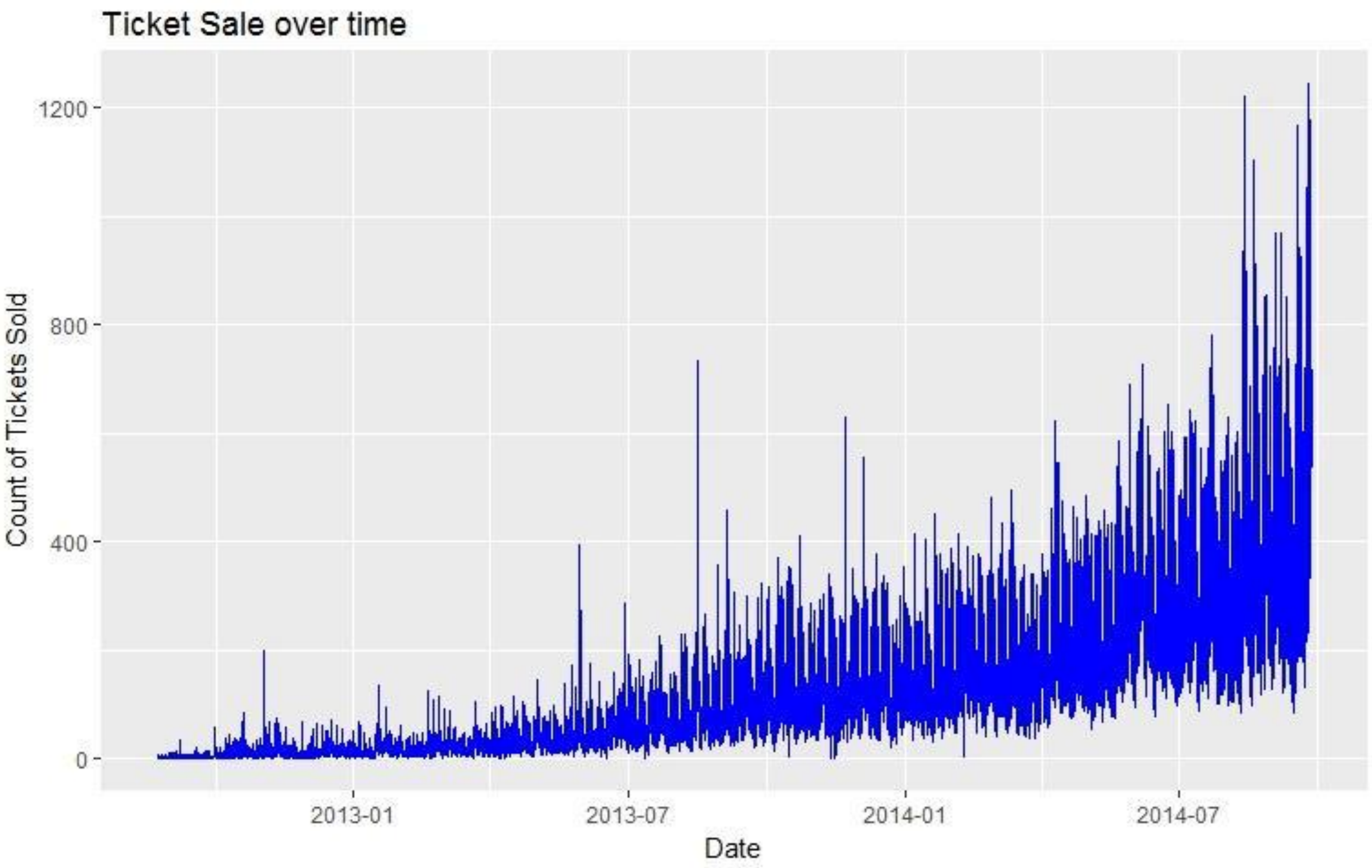


$$\mu = \frac{9 + 9.5 + 11}{3} = 9.833$$

This was trivial, why bother?

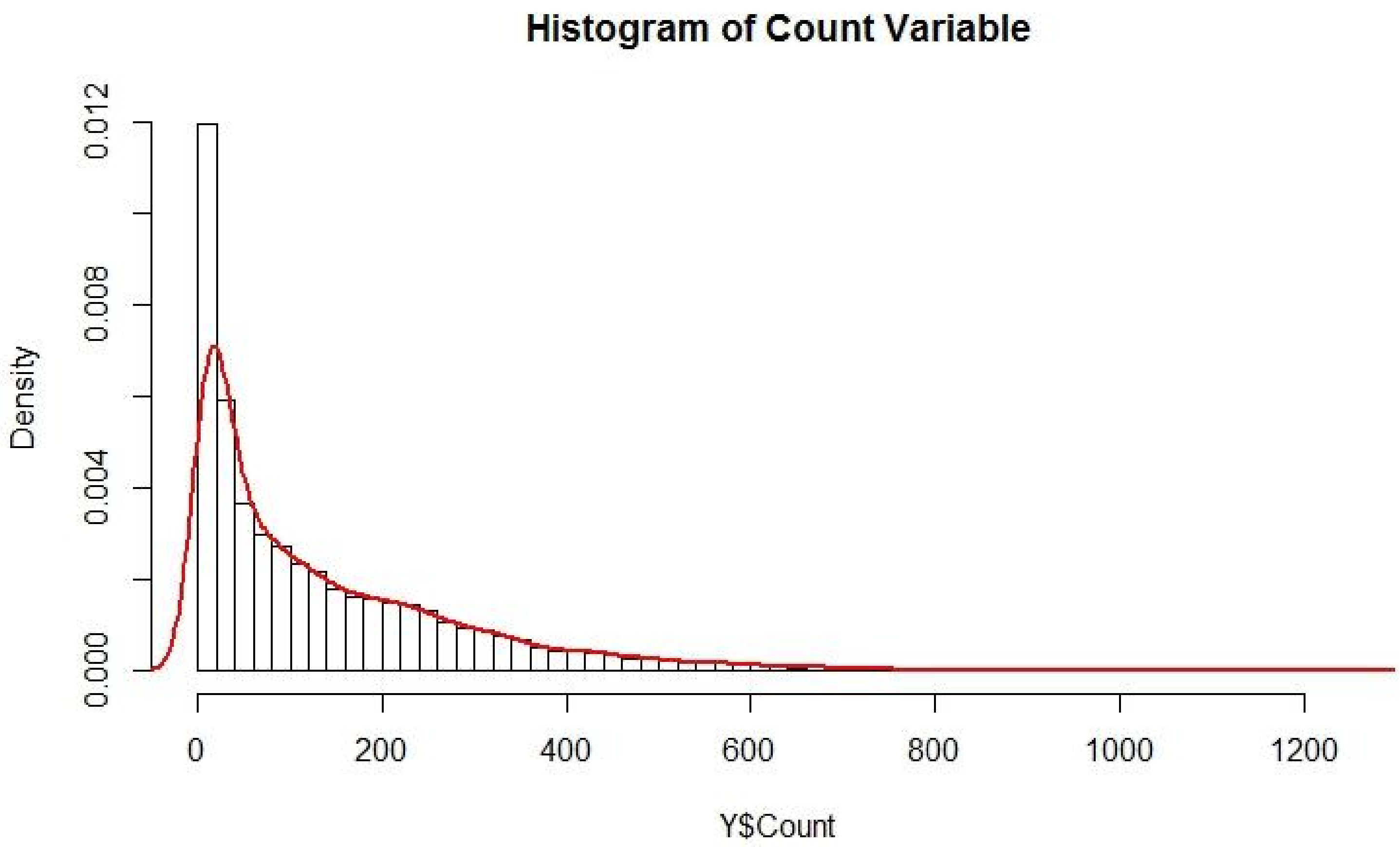


MLE

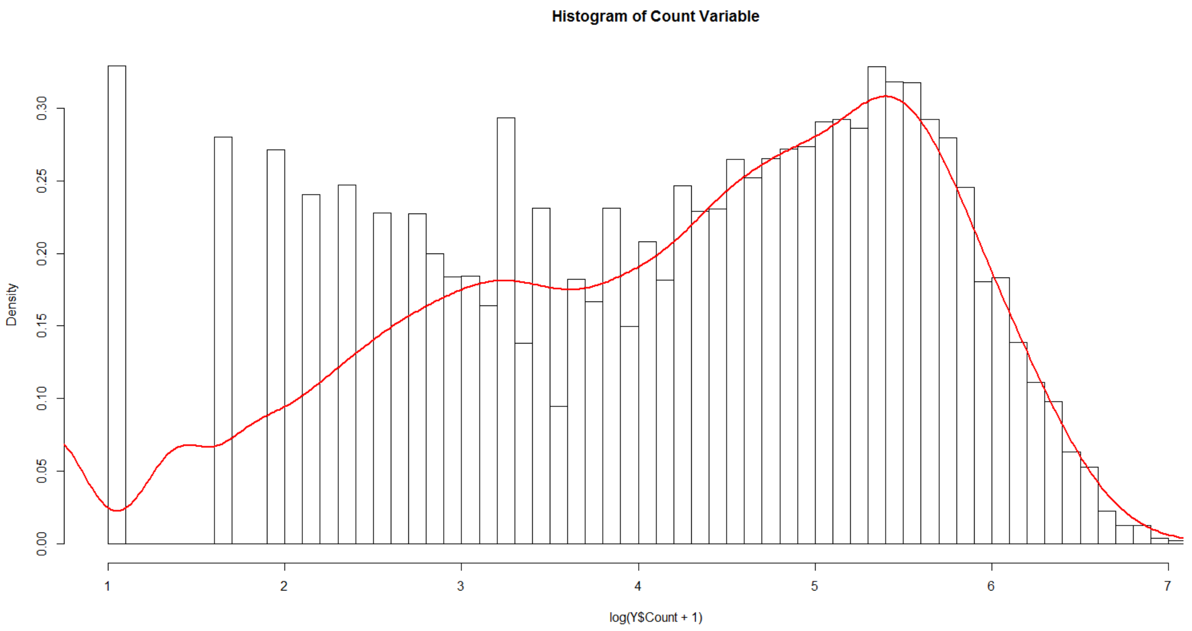


Ticket Sale over time
Count of Tickets Sold
1200
800
400
0
2013-01
2013-07
2014-01
2014-07
Date

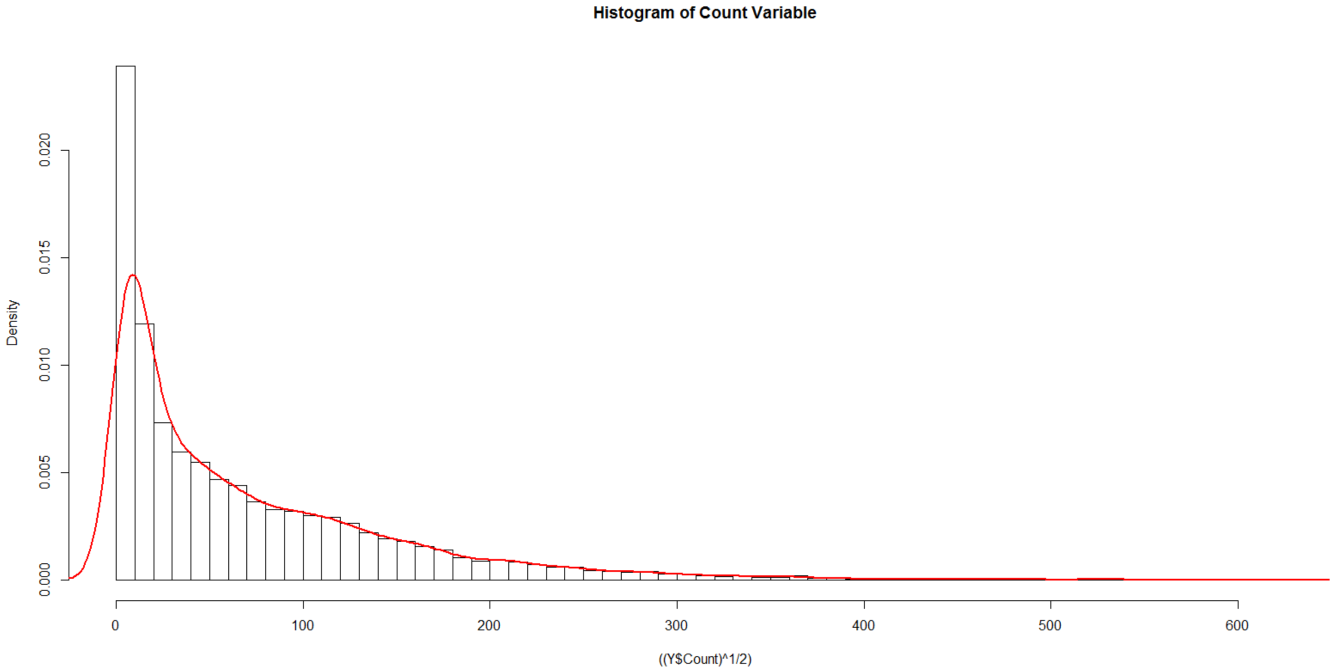
MLE


Histogram of Count Variable
Density
0.000
0.004
0.008
0.012
0
200
400
600
800
1000
1200
Y$Count

MLE

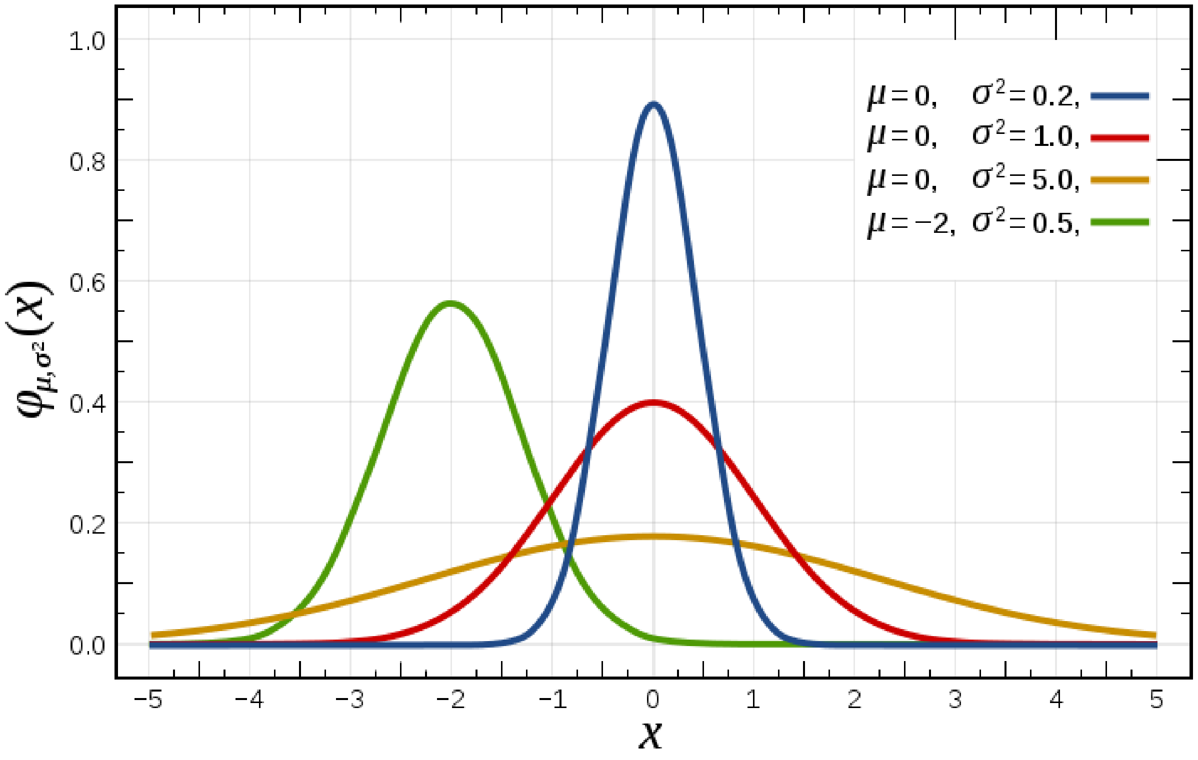


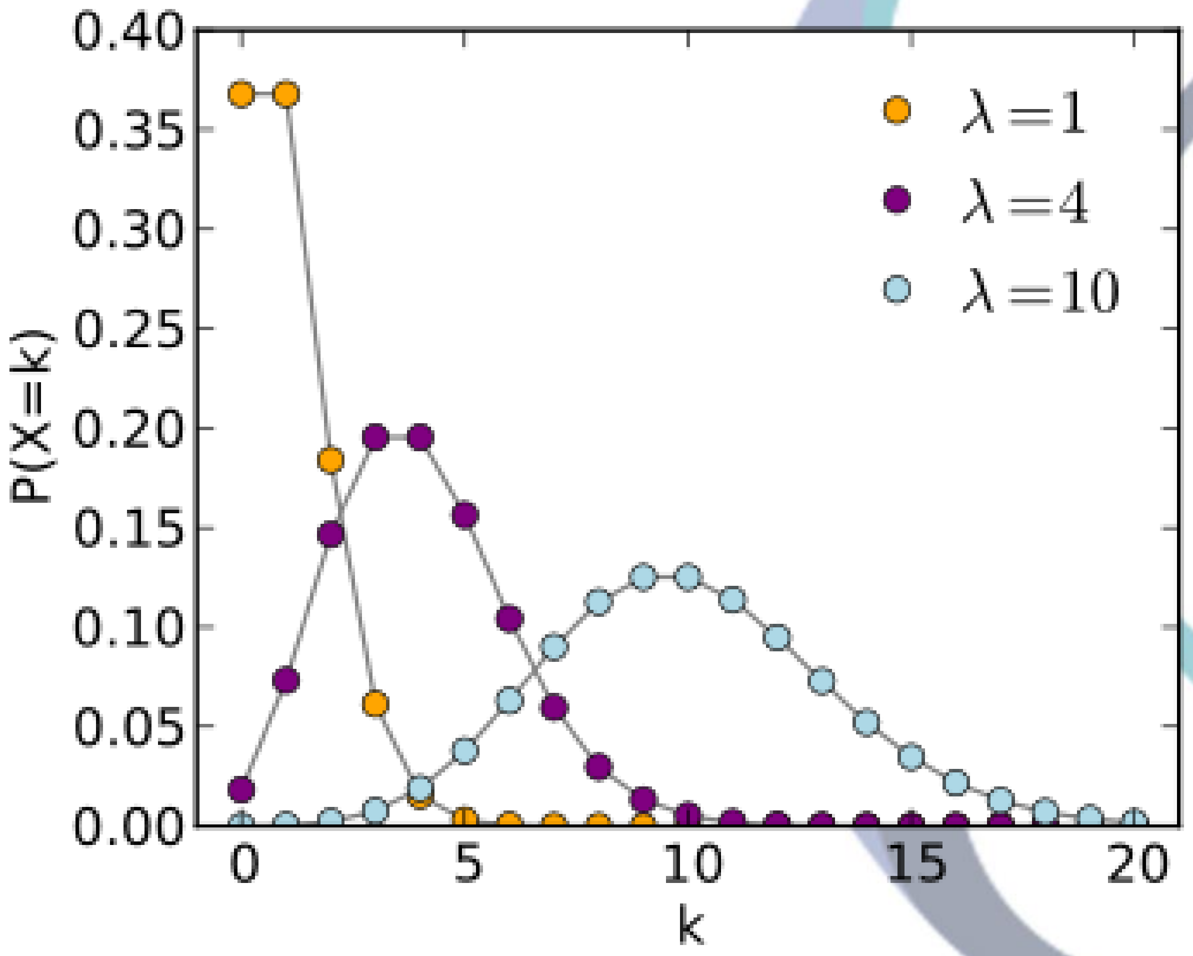
Histogram of Count Variable
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
1
2
3
4
5
6
7
log(Y$Count + 1)



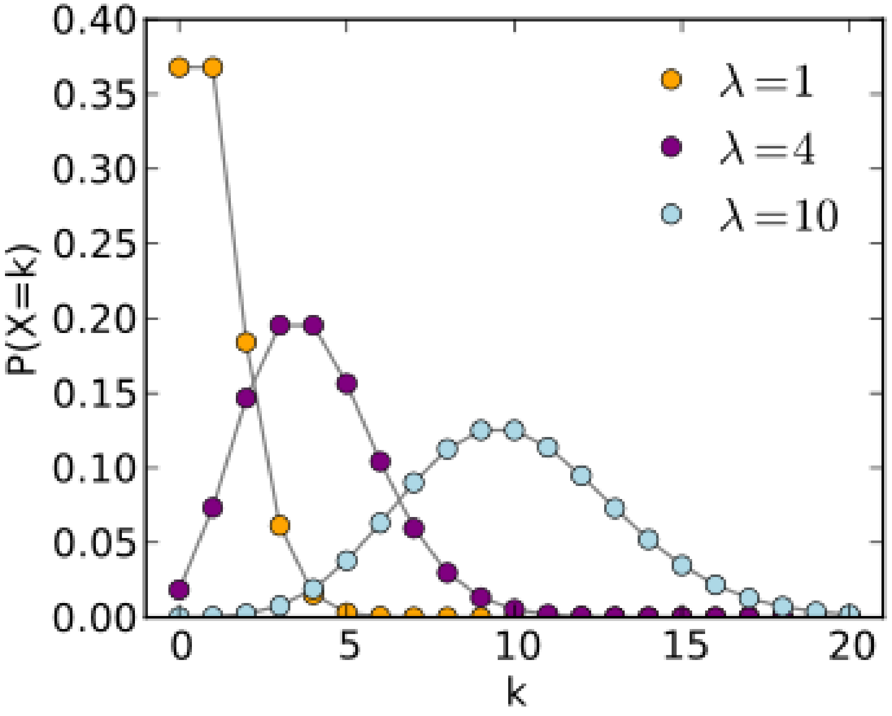
Histogram of Count Variable
Density
0.000
0.005
0.010
0.015
0.020
0
100
200
300
400
500
600
((Y$Count)^1/2)

What distribution?





What distribution?



$$\log(\mu_i) = x_i' \theta$$

or

$$\mu_i = \exp(x_i' \theta)$$

$$\Pr\{Y = y \mid \mu\} = (e^{-\mu} \mu^y)/ y!$$

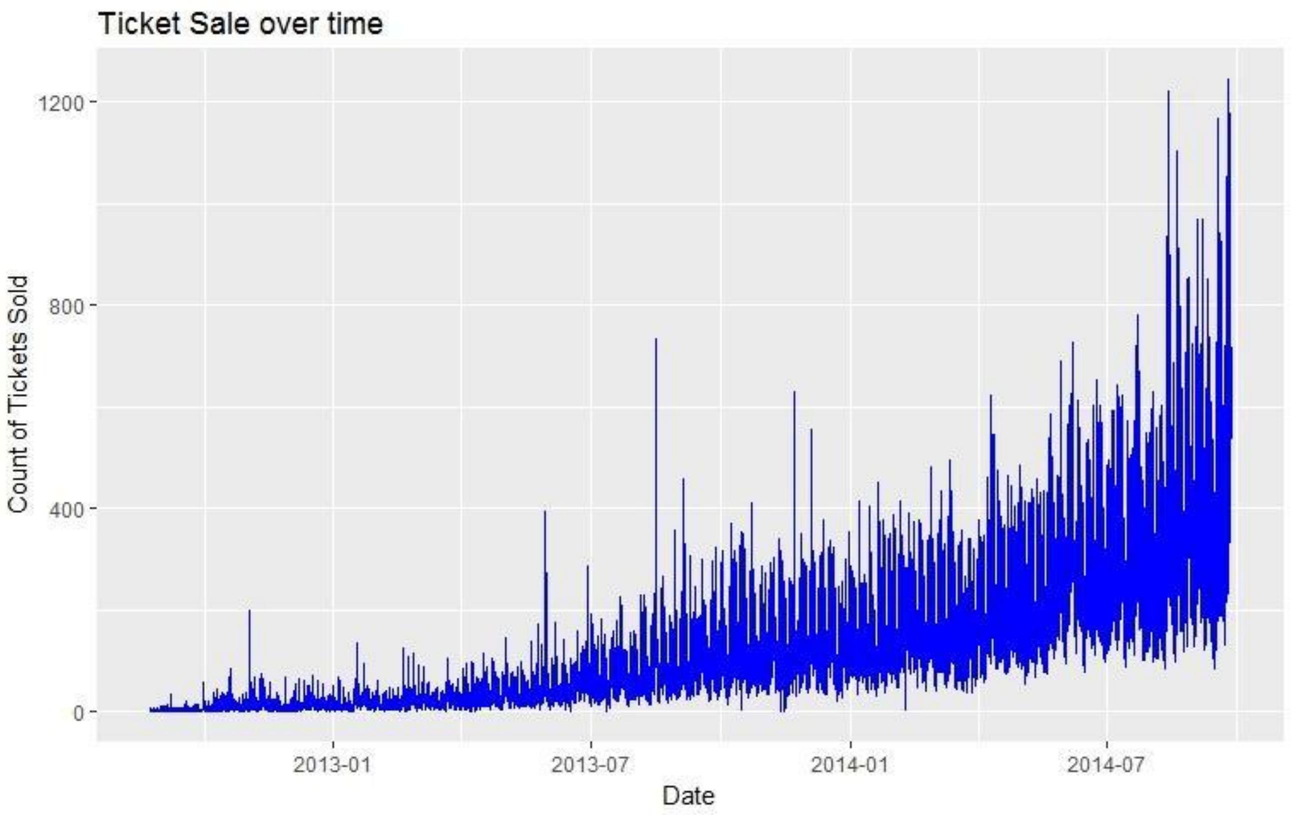
$$LL(\theta) = \sum \{y_i \log(\mu_i) - \mu_i\}, \qquad \text{-------- Eq. 1}$$

What distribution?

$$Pr\{Y = y | \mu\} = (e^{-\mu} \mu^{y})/ y!$$

$$\mu = \exp(\theta_0 + age * \theta_1)$$

$$LL(\theta) = \sum \{y_i.(\theta_0 + age * \theta_1) - \exp(\theta_0 + age * \theta_1)\}$$

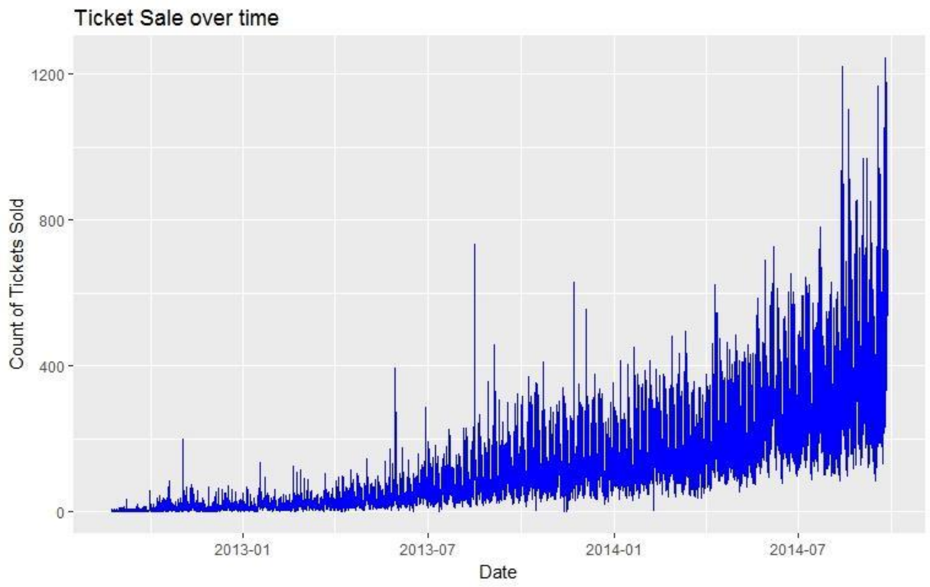


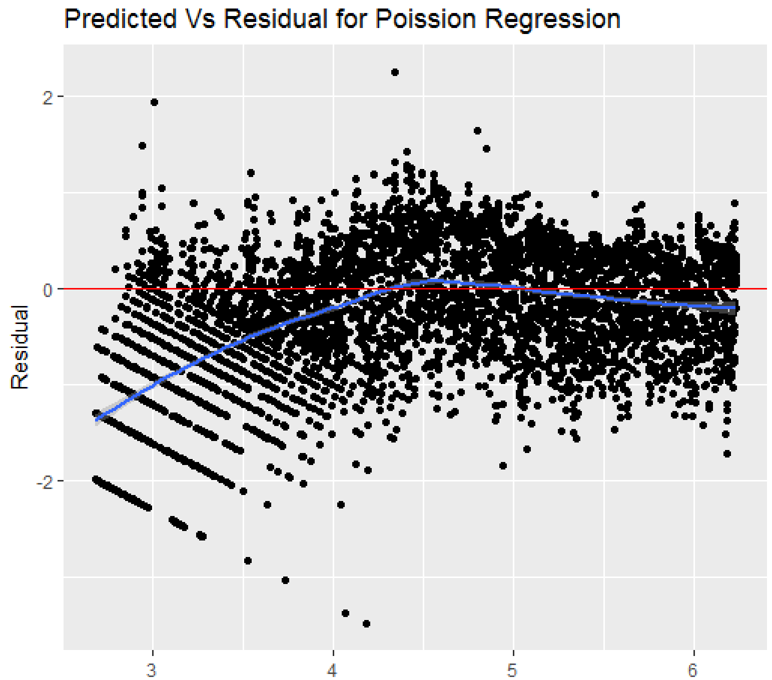
What distribution?

$$\Pr\{Y = y \mid \mu\} = (e^{-\mu}\mu^{y})/\,y!$$

$$\mu = \exp(\theta_0 + age * \theta_1)$$

$$LL(\theta) = \sum \{y_i.(\theta_0 + age * \theta_{1)} - \exp(\theta_0 + age * \theta_{1)}\}$$





Ministério da Ciência, Tecnologia e Inovações

Centro Brasileiro de Pesquisas Físicas

Métodos para Análise de grande volume de dados e Astroinformática

Clécio Roque De Bom – debom@cbpf.br

clearnightsrthebest.com