The Binomial Distribution

N: number of trials

p: probability of success on each trial

K: number of successes

Probability mass function

$$f(k; N, p) = \binom{N}{K} p^{K} (1-p)^{N-K}$$

$\binom{N}{K}$ is the binomial coefficient $\frac{N!}{K!(N-K)!}$

X is a binomial Distributed random variable

$E[X] = Np$

$Var[X] = Np(1-p)$

The Poisson Distribution

λ: Expected counts λ = Rate * Observation Time

K: observed counts

Probability Mass function

$$f(K;\lambda) = \frac{\lambda^K e^{-\lambda}}{K!}$$

Arises when counting events that occur at a constant rate given a specific observation time.

A limit of the Binomial Distribution

when $\lambda = Np$

$N \to \infty$

X is Poison Distributed Random variable

$E[X] = \lambda$

$Var[X] = \lambda$

Notice for Binomial Dist.

$$Var[X] = Np(1-p)$$

as $N \to \infty$ with Np fixed at λ

$p \to 0$

$$Np(1-p) \to \lambda$$

The Gaussian (Normal) Distribution

μ: Expected (mean) value

σ^2: Variance

Probability Density Function (PDF)

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

X is a Gaussian Distributed Random Variable

$E[X] = \mu$

$Var[x] = \sigma^2$

The Distribution of the sum of random variables tend towards a Gaussian distribution.

Random Variable

A random variable, X, is a variable whose possible values are the outcomes of a random phenomenon.

For the Discrete case, the probability of X to take any particular value is given by its PMF. $P(X=x) = f(x)$

For the continuous case, the probability of X to be found in an interval x to x+dx is the PDF.

Expected Value

$E[X]$ is the mean, or first moment of the random variable X.

Discrete Distribution

$$E[X] = \sum_{i=1}^{\infty} x_i P_i$$ where p_i is the probability of x_i

Continuous Distribution

$$E[x] = \int_{-\infty}^{\infty} x f(x)\,dx$$ where $f(x)$ is the probability distribution function.

Variance

$$Var[X] = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$

Discrete Dist.

$$Var[X] = \sum_{i=1}^{\infty} p_i (x_i - \mu)^2$$

Continuous Dist

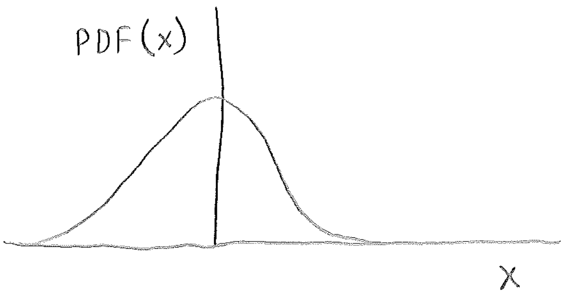
$$Var[X] = \int (x-\mu)^2 f(x)\,dx$$

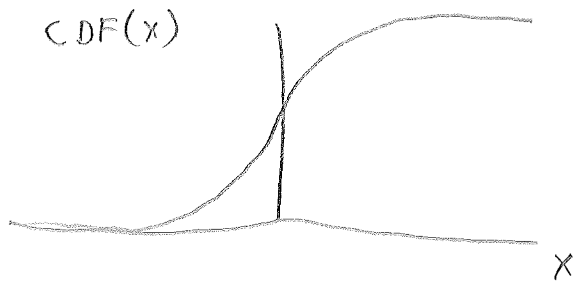
Generating Random Numbers from a Known PDF or PMF

Inversion Method

$$CPF(x) = \int_{-\infty}^{x} PDF(x')\,dx'$$

CDF: cumulative distribution function





Y has a uniform distribution on $[0,1]$

$X = CDF^{-1}(Y)$

X is distributed as $PDF(x)$

Example: Normal Distribution

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

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt$$

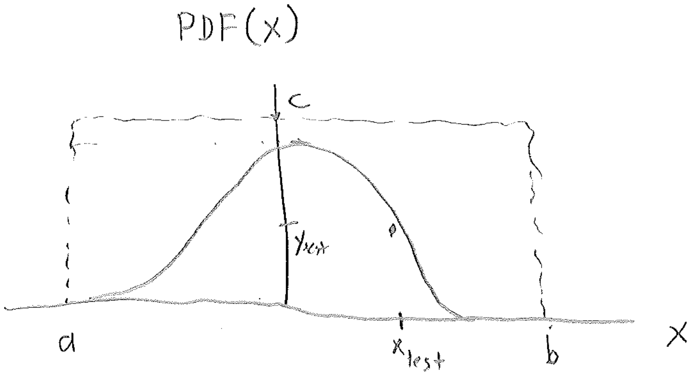
$$CDF(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$CDF^{-1}(x) = \mu + \sigma\sqrt{2}\,\mathrm{erf}^{-1}[2x-1]$$

see matlab

GenRandomNumbers.m

Accept - Reject Method



$c \geq$ Maximum of $PDF(x)$

(1.) x_{test} is selected from uniform distribution on $[a, b]$

(2.) y_{test} is selected from uniform distribution on $[0, c]$

(3.) Accept x_{test} if $y_{test} \leq PDF(x_{test})$

The Central Limit Theorem

$X_1, X_2 \ldots$ are independent, identically distributed random variables having mean μ, and finite non-zero variance σ^2

$S_n = X_1 + X_2 + \ldots X_n$

$$Z_n = \left(\frac{S_n - n\mu}{\sigma \sqrt{n}} \right)$$

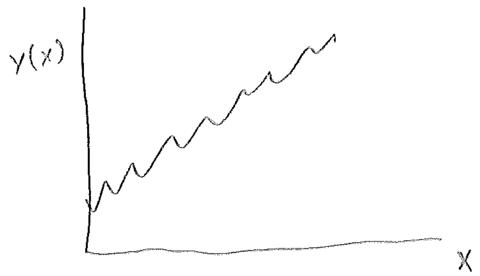
$$Z_n \rightarrow \mathcal{N}(0,1) \quad \text{as} \quad n \rightarrow \infty$$

$\mathcal{N}(\mu, \sigma^2)$ is a Normal distribution with mean μ, variance σ^2.

See CLT Example.m

Note $\ln L(\theta|\vec{x}) = \ln \prod_{n=1}^{N} P(x_n|\theta) = \sum_{n=1}^{N} \ln P(x_n|\theta)$

Example:



$\theta = \begin{pmatrix} a \\ b \end{pmatrix}$

$y(x)$: observed values.

Expected values at x are $\mu(x) = a + bx$

observed values are normally dist. around μ with variance σ^2

$$L(\theta|\vec{y}) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\mu(x_n) - y(x_n))^2/2\sigma^2}$$

$$\ln L(\theta|\vec{y}) = \sum_{n=1}^{N} \left[-\frac{1}{2\sigma^2} (\mu(x_n) - y(x_n))^2 - \ln(2\pi\sigma^2) \right]$$

finding the max of $\ln L(\theta|\vec{x})$ is the same as finding min of

$$\sum_{n=1}^{N} (\mu(x_n) - y(x_n))^2$$

Least squares fitting is the same as using the MLE when observations are normally distributed!

Estimator: A rule for calculating parameters given observed data,

parameters are often writen as θ,

where θ can be a scalar or vector quantity.

the estimator is writen as $\hat{\theta}$. the estimates are $\hat{\theta}(x)$

small x, not X, which is a Rand. Var.

For Example, Estimate the slope and intercept of a line.

x is a known parameter, $f(x)$ is a measured value,

$$f(x) = a + bx$$

$$\theta = \begin{pmatrix} a \\ b \end{pmatrix}$$

Usually, it is desired to have the minimum variance unbiased estimator (MVUE)

Variance $\quad var(\hat{\theta}(X)) = E\left[\left(\hat{\theta}(X) - E(\hat{\theta})\right)^2\right]$

Bias $\quad B(\hat{\theta}(X)) = E\left[\hat{\theta}(X) - \theta\right]$

unbiased $\quad B(\hat{\theta}(X)) = 0$

Often, the Maximum Likelihood Estimator (MLE) is the MVUE (but not always)

Likelihood

the probability of N independent observations is $\prod_{n=1}^{N} P_n$

for a particular set of observation $\vec{x}$,

the probability of $\vec{x}$ given θ is

$$L(\theta | \vec{x}) = \prod_{n=1}^{N} P(x_n | \theta)$$

where $L(\theta|\vec{x})$ is the likelihood.

This is NOT the probability that θ is correct!

for continuous distributions $L(\theta | \vec{x}) = \prod_{n=1}^{N} f(x_n | \theta)$ (Δx interval dropped as constant factor)

The MLE is

$\hat{\theta}_{MLE} = \arg_{\theta} \max L(\theta | \vec{x})$, meaning finding $\theta = \hat{\theta}_{MLE}$ such that $L(\theta|\vec{x})$ is maximized.

this is identical to

$$\hat{\theta}_{MLE} = \arg_{\theta} \max \ln L(\theta | \vec{x})$$

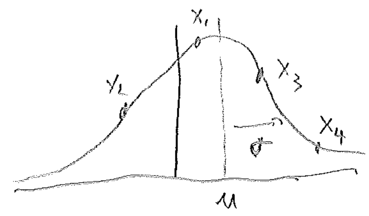
Example: Estimating the mean and variance from a set of observations.

$$L(\theta|\vec{x}) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/2\sigma^2}$$

$$\theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{pmatrix}$$

find where

$$\frac{\partial}{\partial \theta_i} \ln\left[L(\theta|\vec{x})\right] = 0$$



$$\ln\left[L(\theta|\vec{x})\right] = \sum_{n=1}^{N}\left[-\frac{(x_n-\mu)^2}{2\sigma^2} - \frac{1}{2}\ln 2\pi\sigma^2\right]$$

$$\frac{\partial \ln L}{\partial \mu} = \sum_{n=1}^{N} -\frac{(x_n-\hat{\mu})}{\sigma^2} = 0$$

$$\sum_{n=1}^{N} x_n = \sum_{n=1}^{N} \hat{\mu}$$

$$\hat{\mu} = \bar{x} \qquad E[\hat{\mu}] = \mu$$

$$\frac{\partial \ln L}{\partial \sigma} = \sum_{n=1}^{N}\left[\frac{(x_n-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right]$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x})^2$$ ← since the estimated mean is used, $\hat{\sigma}$ is biased

$$\delta_n = \mu - x_n$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(\mu-\delta_n)^2 - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\mu-\delta_i)(\mu-\delta_j)$$

$$E[\delta_i] = 0, \qquad E[\delta_i^2] = \sigma^2$$

$$E[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2$$

the estimator for the unbiased population variance is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{N}(x_n-\bar{x})^2$$

σ^2 population variance

★ - equal only after expectation

s^2 sample variance

$$s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \qquad \sigma^2 = E\left[(x_i-\mu)^2\right] \overset{*}{=} E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$E[s^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n}\left((x_i-\mu)-(\bar{x}-\mu)\right)^2\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 - 2(\bar{x}-\mu)\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu) + \frac{1}{n}\sum_{i=1}^{n}(\bar{x}-\mu)^2\right]$$

$$\overset{*}{=} \sigma^2 + E\left[-2(\bar{x}-\mu)^2 + (\bar{x}-\mu)^2\right]$$

$$= \underbrace{\sigma^2 - E\left[(\bar{x}-\mu)^2\right]}$$

s^2 is a biased estimator for the population mean σ^2

$$E\left[(\bar{x}-\mu)^2\right] = E\left[\left(\frac{1}{n}\sum x_i - \mu\right)^2\right]$$

$$= E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}x_i x_j - \frac{2\mu}{n}\sum_{i=1}^{n}x_i + \mu^2\right]$$

$$\overset{*}{=} E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}x_i x_j - \mu^2\right] \overset{*}{=} E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-\mu)(x_j-\mu)\right]$$

$$= E\left[\frac{1}{n^2}\sum_{i=1}^{n}(x_i-\mu)^2\right] = \frac{1}{n}\sigma^2$$

$$E[S^2] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

$$\sigma^2 = \frac{n}{n-1} E[S^2]$$

so unbiased estimator is

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

$$E\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \sigma^2$$

Ordinary Least Squares (Linear Least Squares)

A model that is linear in its coefficients, θ_j, is given by $y(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + ...$ For a model with M linear coefficients this can be written as

$$y(x) = \sum_{j=1}^{M} \theta_j f_j(x)$$

At x_i this becomes

$$y(x_i) = \sum_{j=1}^{M} \theta_j f_j(x_i)$$

There are N data points corresponding to observations made at $x_1 ... x_N$, so lets define a 'design matrix' $\mathbf{X}$ of size $N \times M$ such that $X_{i,j} = f_j(x_i)$. Now

$$y(x_i) = \sum_{j=1}^{M} \theta_j X_{i,j}$$

If the data is normally distributed with a known variance of σ^2 then the Likelihood is

$$L(\theta|\vec{y}) = \prod_{i=1}^{N} \frac{1}{1/\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \sum_{j=1}^{M} \theta_j X_{i,j})^2}$$

Finding maximum of this expression (by varying θ) is the same as finding the minimum of the negative, which is equivalent to finding the minimum of the negative of the log of the expression. So look for the minimum of:

$$S(\theta) \equiv \sum_{i=1}^{N} (y_i - \sum_{j}^{M} \theta_j X_{i,j})^2$$

Note that constant terms are dropped since we only care about where the minimum is found, not its value. The minimum is found when for each k,

$$\frac{\partial S}{\partial \theta_k} = 0$$

which written out is:

$$\frac{\partial S}{\partial \theta_k} = \sum_{i=1}^{N} (y_i - \sum_{j=1}^{M} \theta_j X_{i,j}) X_{i,k} = 0$$

Rearranging gives:

$$\sum_{i=1}^{N} y_i X_{i,k} = \sum_{i=1}^{N} \sum_{j=1}^{M} \theta_j X_{i,j} X_{i,k}$$

Since $\mathbf{y}$ is $N \times 1$, and $\mathbf{X}$ is $N \times M$, the left side looks like (kth row of $\mathbf{X}'$) $\times$ $\mathbf{y}$. Since θ is $M \times 1$, the right side looks like (kth row of $\mathbf{X}'$)$\times$ ((jth row of $\mathbf{X}$) $\times$ θ).

After the summations, the only remaining index is the coefficient index. We can write this so it relates two vectors of size $M \times 1$ by writing:

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\theta$$

And finally, to solve for θ,

$$\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Note the 'hat' above θ means the estimated θ.

Fisher Information and the Cramér-Rao Lower Bound

Fisher Information is a measure of the amount of information a random variable X carries about an unknown parameters θ

$$V = \frac{\partial}{\partial\theta} \ln L(\theta|\vec{x}) = \frac{\left(\frac{\partial L(\theta|\vec{x})}{\partial\theta}\right)}{L(\theta|\vec{x})}$$

$\vec{x}$ indicates a set of observations

V is the 'score' and indicates the sensitivity of $L(\theta|\vec{x})$

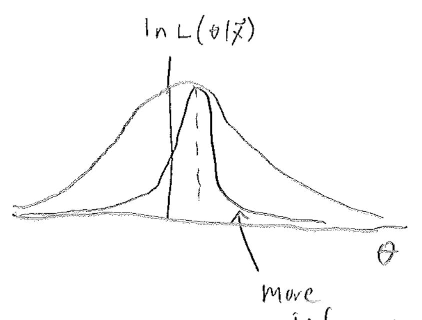
$$E[V] = 0$$

The Fisher Information $I(\theta) = E[V^2] = E\left[\left(\frac{\partial}{\partial\theta} \ln L(\theta|\vec{x})\right)^2\right]$

If:
1. $L(\theta|\vec{x})$ is twice differentiable w.r.t θ
2. $L(\theta|\vec{x})$ has bounded support in x and bounds are independent on θ

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta} \ln L(\theta|\vec{x})\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2} \ln L(\theta|x)\right]$$

↑ note this is the curvature of $\ln L(\theta|x)$



The Cramér Rao Lower Bound states that

$$Var[\hat{\theta}] \geq I(\theta)^{-1}$$

This gives a minimum bound for any unbiased estimator.

An estimator is said to be 'efficient' if it acheives the CRLB. Often, the MLE is efficient.

For $\theta = \{\theta_1, \theta_2 .. \theta_N\}^T$ (multiple parameters)

V is a vector $V = \left\{ \frac{\partial \ln L(\theta|\vec{x})}{\partial \theta_1}, \dots \frac{\partial \ln L(\theta|\vec{x})}{\partial \theta_N} \right\}$

and $I(\theta)$ is the Fisher Information matrix

$$I(\theta) = E\left[V V^T\right]$$

$$I(\theta)_{ij} = E\left[\left(\frac{\partial \ln L(\theta|\vec{x})}{\partial \theta_i}\right)\left(\frac{\partial \ln L(\theta|\vec{x})}{\partial \theta_j}\right)\right]$$

$$= -E\left[\frac{\partial^2 \ln L(\theta|\vec{x})}{\partial \theta_i \partial \theta_j}\right]$$

and the CRLB is $cov(\hat{\theta}) \geq I(\theta)^{-1}$

CRLB Example: Fitting a line using Least Squares

$$y(x_n) = A + B x_n + w_n$$ indicates Gaussian noise, σ known

$$\mu(x_n) = A + B x_n$$

$$\theta = \{A \quad B\}^T$$

$$L(\theta | \vec{y}) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_n - \mu_n)^2/2\sigma^2}$$

$$\frac{\partial \ln L(\theta|\vec{y})}{\partial A} = \sum_{n=1}^{N} \frac{\partial}{\partial A}\left[-\frac{1}{2\sigma^2}\left(y_n - (A + Bx_n)\right)^2\right] = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(y_n - A - Bx_n\right)$$

$$\frac{\partial \ln L(\theta|\vec{y})}{\partial B} = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(y_n - A - Bx_n\right)x_n$$

$$\frac{\partial \ln L}{\partial A^2} = -\frac{N}{\sigma^2}$$

$$\frac{\partial \ln L}{\partial B^2} = -\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n^2$$

$$\frac{\partial \ln L}{\partial A \partial B} = -\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n$$

does not depend on y, the observable

$$I(\theta)_{11} = -E\left[-\frac{N}{\sigma^2}\right] = \frac{N}{\sigma^2}$$

$$I(\theta)_{12} = I(\theta)_{21} = -E\left[-\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n\right] = \frac{1}{\sigma^2}\sum_{n=1}^{N} x_n$$

$$I(\theta)_{22} = -E\left[-\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n^2\right] = \frac{1}{\sigma^2}\sum_{n=1}^{N} x_n^2$$

$$I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} N & \sum_{n=1}^{N} x_n \\ \sum_{n=1}^{N} x_n & \sum_{n=1}^{N} x_n^2 \end{bmatrix}$$

$$\mathrm{var}(\hat{A}) \geq \left(I(\theta)^{-1}\right)_{11}$$

$$\mathrm{var}(\hat{B}) \geq \left(I(\theta)^{-1}\right)_{22}$$

See matlab Example.

Assingment of Uncertainty to Estimated parameters

Standard error of the mean

Implicit assumption of a Normal population

s : sample standard deviation

σ: population std. dev.

unbiased estimate of σ is $s\sqrt{\frac{N}{N-1}}$

Standard error of the mean is

$$SE_{\bar{x}} \cong \frac{s}{\sqrt{N}}$$ with bias correction

$$SE_{\bar{x}} = \frac{s}{\sqrt{N}}\sqrt{\frac{N}{N-1}}$$

The mean can be reported as

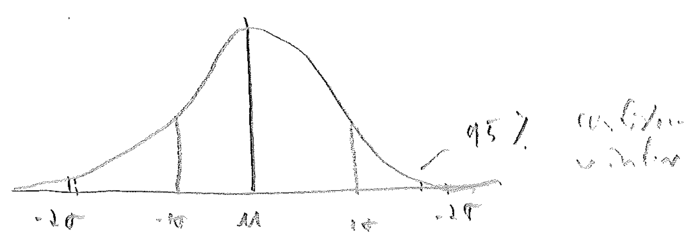
$\bar{x} \pm SE_{\bar{x}}$, where the error is explicitly stated as the standard error of the mean.

The mean can also be reported with a 'confidence interval'

If $\bar{x}$ is the true population mean, e: 95% of observations would fall into the 95% confidence interval.

for 95% confidence interval

$\bar{x} \pm SE_{\bar{x}} \cdot 1.96$



Standard Error can also be given as the standard deviation of the estimator $\hat{\theta}(x)$.

If the estimator is 'efficient' this can be calculated from the CRLB.

$$\theta_i \pm \left[\left(I(\theta)^{-1} \right)_{ii} \right]^{1/2}$$

Properties of the estimator could also be found from 'Monte-Carlo' simulations

The standard error can also be given by estimating the variance of the estimator using a given data set.

'Hessian' matrix

$$H(\ln L(\theta|\vec{y}))_{ij} = \frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}$$

$$SE \approx \left[\left[H(\ln L(\theta|\vec{y})) \right]^{-1}_{ii} \right]^{1/2}$$

Always clearly describe how a parameter's uncertainty is calculated!

the χ^2 distribution

$$Q = \sum_{i=1}^{K} x_i^2 \quad , \quad x_i: \text{ independent, normally distributed with } \mu = 0, \ \sigma^2 = 1$$

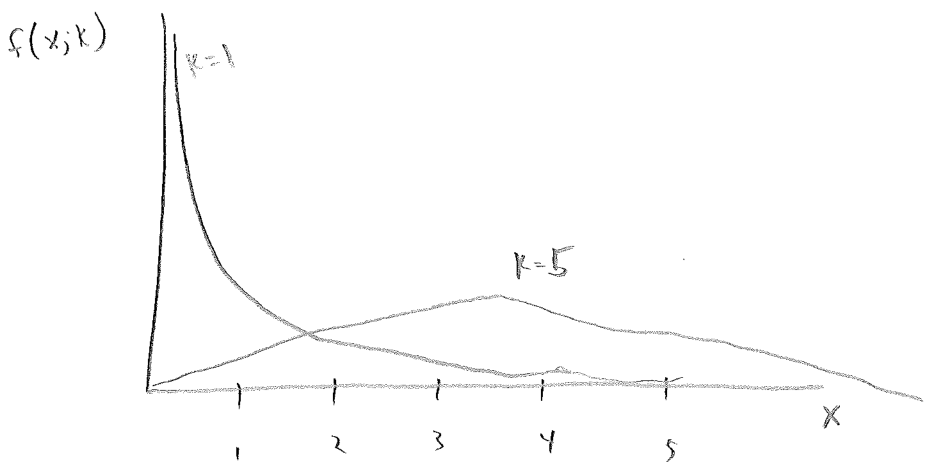
Q is "χ^2" distributed with K degrees of freedom

the χ^2 distribution:

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2 - 1} e^{-x/2} \qquad \text{(PDF)}$$

Mean: K

variance: 2K



Hypothesis Testing

$\vec{y}$: set of N observations

θ: set of M parameters

w_n: independent MLE of each observation y_n
(typically $w_n = y_n$)

$$D = -2 \ln \frac{L(\hat{\theta} | \vec{y})}{L(w | \vec{y})}$$

D is asymptotically χ^2 distributed as $N \to \infty$,

with $K = N - M$ degress of freedom.

Since we know how D is distributed, we

can calculate a p-value:

The probability, Assuming Null hypothesis is true, of obtaining D that is at least as extreme as that observed.

Test the Null Hypothesis; which is usually:

H_0: The model, parameritized by $\hat{\theta}$, is correct.

$$p = 1 - \int_0^D f(x;K)\, dx$$

if $p > 0.01$, we accept H_0 at the 0.01 significance level.

Does not prove H_0 is true!

See X2TestExample.m

Hypothesis test of a restricted model

$\hat{\theta}$: unrestricted

$\hat{\theta}'$: restricted

(ie $\hat{\theta} = \{\theta_1, \theta_2, \theta_3\}^T$

$\hat{\theta}' = \{\theta_1', \theta_2', 0\}$)

k: parameters in unrestricted model

k': number of restricted parameters

$$D = -2\ln\left(\frac{L(\hat{\theta}' \mid \vec{y})}{L(\hat{\theta} \mid \vec{y})}\right)$$

is asymptotically χ^2 distributed

with k' degrees of freedom.

Example: Is background offset term needed?

unrestricted Expectation model

$$\mu(x) = \theta_1 e^{-\theta_2 x} + \theta_3$$

$$\theta = \{\theta_1, \theta_2, \theta_3\}^T$$

restricted model

$$\theta = \{\theta_1, \theta_2, 0\}^T$$

D is χ^2 dist. with 1 degree of freedom

Null hypothesis: there is no background term

$\chi^2 = 6.64$ when $p = 0.01$ for 1 degree of

freedom, Do not reject Ho

if D is < 6.64

Covariance Matrix

X_1, X_2 are random variables with finite variance

elements of the covariance matrix

are $$V_{ij} = E\left[\left(X_i - E[X_i]\right)\left(X_2 - E[X_2]\right)\right] = \langle (x_1 - \langle x_1 \rangle)(x_2 - \langle x_2 \rangle) \rangle$$

$$V_{ii} = \sigma_{x_i}^2$$
$$V_{ij} = \mathrm{cov}(x_i, x_j)$$

use in Error propogation

$q(x, y)$ is a function of the measured

values x, y

$$\langle x \rangle = \int P(x,y)\, x\, dx$$
$$\langle q \rangle = \int P(x,y)\, q(x,y)\, dx\, dy$$

Expand q around $\langle x \rangle, \langle y \rangle$

$$q(x,y) \cong q(\langle x \rangle, \langle y \rangle) + \frac{\partial q}{\partial x}\Big|(x - \langle x \rangle) + \frac{\partial q}{\partial y}\Big|(y - \langle y \rangle)$$

where $\frac{\partial q}{\partial x}\Big|$, $\frac{\partial q}{\partial y}\Big|$ are evaluated at $\langle x \rangle, \langle y \rangle$

and therefore not a funtion of x, y.

Then $\langle q \rangle = q(\langle x \rangle, \langle y \rangle)$

$$\langle (x - \langle x \rangle) \rangle = 0 = \langle y - \langle y \rangle \rangle$$

$$\sigma_q^2 = \langle (q - \langle q \rangle)^2 \rangle$$

$$= \left\langle \left(\frac{\partial q}{\partial x}\Big| (x - \langle x \rangle) + \frac{\partial q}{\partial y}\Big| (y - \langle y \rangle) \right)^2 \right\rangle$$

$$= \left\langle \left(\frac{\partial q}{\partial x}\Big|\right)^2 (x - \langle x \rangle)^2 + \left(\frac{\partial q}{\partial y}\Big|\right)^2 (y - \langle y \rangle)^2 + \left(\frac{\partial q}{\partial x}\Big| \frac{\partial q}{\partial y}\Big|\right) (x - \langle x \rangle)(y - \langle y \rangle) \right\rangle$$

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\Big|\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\Big|\right)^2 \sigma_y^2 + 2 \left(\frac{\partial q}{\partial x}\Big|\right)\left(\frac{\partial q}{\partial y}\Big|\right) \mathrm{cov}(x, y)$$

Note that this can be written as

$$\sigma_q^2 = \begin{pmatrix} \frac{\partial q}{\partial x}\Big| & \frac{\partial q}{\partial y}\Big| \end{pmatrix} V \begin{pmatrix} \frac{\partial q}{\partial x}\Big| \\ \frac{\partial q}{\partial y}\Big| \end{pmatrix}$$

and for unbiased, efficient estimates

$$V \to [I(\theta)]^{-1}.$$

If x, y are uncorrelated $P(x, y) = P(x) P(y)$

$$\mathrm{cov}(x, y) = \int dx\, P(x) (x - \langle x \rangle) \int dy\, P(y) (y - \langle y \rangle) = 0$$

we can define a correlation coefficient as

$$r = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}$$

In general, we can expand q to second order in a Taylor series expansion:

$$q(\beta, b) \cong q(\langle\beta\rangle, \langle b\rangle) + \frac{\partial q}{\partial \beta}\Big| (\beta - \langle\beta\rangle) + \frac{\partial q}{\partial b}\Big| (b - \langle b\rangle)$$

$$+ \frac{1}{2}\Big[\frac{\partial^2 q}{\partial \beta^2}\Big| (\beta - \langle\beta\rangle)^2 + \frac{\partial^2 q}{\partial b^2}\Big| (b - \langle b\rangle)^2$$

$$+ 2\left(\frac{\partial q}{\partial b}\Big|\right)\left(\frac{\partial q}{\partial \beta}\Big|\right)(b - \langle b\rangle)(\beta - \langle\beta\rangle)\Big]$$

$$\langle q(\beta, b)\rangle \cong q(\langle\beta\rangle, \langle b\rangle) + \frac{1}{2}\frac{\partial^2 q}{\partial \beta^2}\Big| \sigma_\beta^2 + \frac{1}{2}\frac{\partial^2 q}{\partial b^2}\Big| \sigma_b^2 + \left(\frac{\partial q}{\partial b}\Big|\right)\left(\frac{\partial q}{\partial \beta}\Big|\right) \mathrm{cov}(b, \beta)$$

In our Example,

$$q(\beta, b) = T = \frac{1}{\beta b} \qquad \mathrm{cov}(b, \beta) = 0$$

$$\frac{\partial^2 q}{\partial \beta^2} = \frac{2}{\beta^3 b}$$

$$\frac{\partial^2 q}{\partial b} = \frac{2}{\beta b^3}$$

$$\langle q(\beta, b)\rangle \cong \frac{1}{\beta b} + \frac{1}{\beta^3 b}\sigma_\beta^2 + \frac{1}{\beta b^3}\sigma_b^2$$

An error budget example

$$\tau = \frac{1}{b\beta} \qquad b \sim 2 \qquad \beta \sim 3 \qquad \tau \sim 0.17$$

	error in parameter	Transfer function	Error in τ
β	.3	$\frac{1}{\beta^2 b}$	.017
b	.2	$\frac{1}{\beta b^2}$	.017
Total RSS error			.024

Root Sum of Squares (RSS)

$$\frac{\partial \tau}{\partial \beta} = -\frac{1}{\beta^2 b} \qquad \frac{\partial \tau}{\partial b} = -\frac{1}{\beta b^2}$$

RSS error assumes the errors are uncorrelated and represent the standard error of the parameters.

Expectation model

$$\mu(x) = Ae^{-\beta x} + B$$

x: channel number

$$x = a + bt$$

t: time at center of bin

$$\tau = \frac{1}{\beta b}$$

$$\mu(t) = Ae^{-\beta(a+bt)} + B$$

$$= Ae^{-\beta bt} e^{-\beta a} + B$$

$N_0 = Ae^{-\beta a}$

Background $= B$

$\tau = \frac{1}{\beta b}$

If we want τ with 1% error,

$$\sigma_\tau = \frac{\tau}{100}$$

How well do we need to know

β and b? Give in relative error.

$$\tau(\beta, b) = \frac{1}{\beta b}$$

$$\sigma_\tau^2 = \begin{pmatrix} \frac{\partial \tau}{\partial \beta} & \frac{\partial \tau}{\partial b} \end{pmatrix} \begin{bmatrix} \sigma_\beta^2 & \mathrm{cov}(\beta, b) \\ \mathrm{cov}(\beta, b) & \sigma_b^2 \end{bmatrix} \begin{bmatrix} \frac{\partial \tau}{\partial \beta} \\ \frac{\partial \tau}{\partial b} \end{bmatrix}$$

in this case β and b are measured in independent measurements and $\mathrm{cov}(\beta, b) = 0$

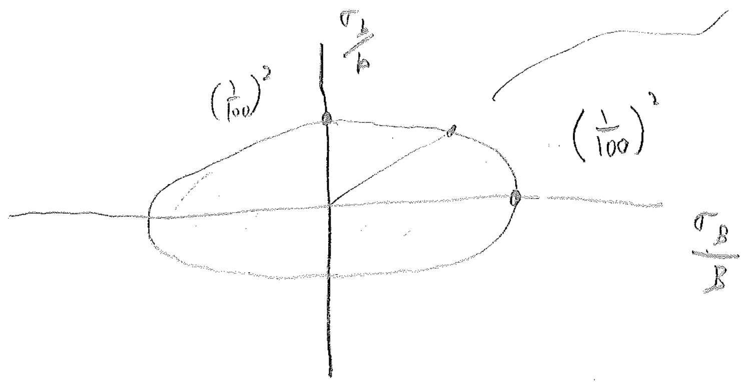
$$\sigma_\tau^2 = \left(\frac{\partial \tau}{\partial \beta}\right)^2 \sigma_\beta^2 + \left(\frac{\partial \tau}{\partial b}\right)^2 \sigma_b^2$$

$$\frac{\partial \tau}{\partial \beta} = -\frac{1}{\beta^2 b} \qquad \frac{\partial \tau}{\partial b} = -\frac{1}{\beta b^2}$$

$$\left(\frac{\sigma_T}{T}\right)^2 = \left(\frac{1}{100}\right)^2 \geq \left(\frac{\sigma_B}{B}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2$$

$$\frac{\left(\frac{\sigma_B}{B}\right)^2}{\left(\frac{1}{100}\right)^2} + \frac{\left(\frac{\sigma_b}{b}\right)^2}{\left(\frac{1}{100}\right)^2} \leq 1$$

$$\frac{\sigma_b}{b} \approx \frac{\sigma_B}{B} = \frac{1}{\sqrt{2}\cdot 100}$$



see ErrorProp1.m

Second order Bias correction

Note: $\left\langle \frac{1}{Bb} \right\rangle \neq \frac{1}{\langle B \rangle \langle b \rangle}$ ↑ what we are calculating

↗ what we want

$$q(B,b) = \frac{1}{Bb}$$

$$\langle q \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(B,b)\, q(B,b)\, dB\, db$$

in this case $P(B,b) = P(B)\,P(b)$

$$\langle q \rangle = \int_{-\infty}^{\infty} P(B)\frac{1}{B}\, dB \int_{-\infty}^{\infty} P(b)\frac{1}{b}\, db$$

for $P(B) = \mathcal{N}(B, \sigma_B^2)$

there is not a closed form expression for $\left\langle \frac{1}{B} \right\rangle$

The Joint Normal Distribution

K: dimensionality

$\mu \in R^K$

$\Sigma \in R^{K \times K}$

$x \in R^K$

X is a K-dimensional random vector $X = [x_1, x_2, \dots x_K]$

$$P(x) = \frac{1}{2\pi^{(K/2)} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)}$$

$|\Sigma|$ is the determinant of Σ

Example for $K=2$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & cov(x_1, x_2) \\ cov(x_1, x_2) & \sigma_2^2 \end{pmatrix}$$ ← variance, covariance matrix

$$r \equiv \frac{cov(x_1, x_2)}{\sigma_1 \sigma_2}$$ ← covariance coefficient

$$\Sigma = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$|\Sigma| = \sigma_1^2\sigma_2^2 - r^2\sigma_1^2\sigma_2^2$$

$$|\Sigma|^{1/2} = \sigma_1\sigma_2\sqrt{1-r}$$

find Inverse of Σ

in general $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $\quad M^{-1}M = I$

$$M = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - r^2\sigma_1^2\sigma_2^2}\begin{bmatrix} \sigma_2^2 & -r\sigma_1\sigma_2 \\ -r\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1-r^2}\begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-r}{\sigma_1\sigma_2} \\ \frac{-r}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}$$

lets take $\mu_1 = \mu_2 = 0$ $\quad X_1 \to x \quad \sigma_1 \to \sigma_x$

$X_2 \to y \quad \sigma_2 \to \sigma_y$

$$P(x, y) = \frac{1}{2\pi\sigma_x\sigma_y}\frac{1}{\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2rxy}{\sigma_x\sigma_y}\right]}$$

Note as $r = 0$

$$P(x,y) = \mathcal{N}(\mu_0 \sigma_x^2)\mathcal{N}(\mu_0 \sigma_y^2)$$

MLE of Σ is

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$$

This is biased, just like in 1D

Un biased sample covariance

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

$x \in R^k$

Using covariance matrix to give predictive interval

Expectation model

$$\mu(x_n) = A + Bx_n$$

observed data

$$y(x_n) = A + Bx_n + w_n$$ ← Gaussian noise with known sigma.

$\theta = \{B \; A\}^T$ ← (B, A switched to match polyfit)

$$L(\theta \mid y) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_n - \mu_n)^2/2\sigma^2}$$

see CRLB example

⋮

$$I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=1}^{N} x_n^2 & \sum_{n=1}^{N} x_n \\ \sum_{n=1}^{N} x_n & N \end{bmatrix}$$

covariance matrix is

$$V \equiv \Sigma = I(\theta)^{-1}$$

calculate expected value and interval that will contain ~68% (1σ) of future estimates,

for any x (x does not have to be one of the x_n)

$$\hat{y} = \hat{A} + \hat{B}x$$

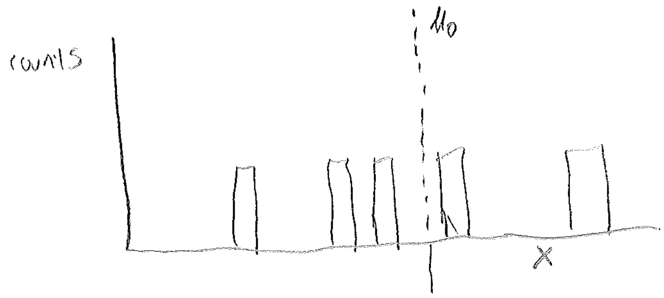
$$q(A,B) = A + Bx$$

$$\sigma_y^2 = \left(\frac{\partial q}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial q}{\partial B}\right)^2 \sigma_B^2 + 2\left(\frac{\partial q}{\partial A}\right)\left(\frac{\partial q}{\partial B}\right) \mathrm{cov}(A,B)$$

$$= \sigma_A^2 + x^2\sigma_B^2 + 2x\,\mathrm{cov}(A,B)$$

see PredictiveInterval.m

Student t-test



Is the sample consistent with population mean of μ_0?

Null Hypothesis: μ_0 is the population mean.

make a test statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

s is the sample standard deviation

$$s = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

t is distributed with a

student t-distribution with $\nu = n-1$
degrees of freedom.

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

a p-value can be calculated and compared
to a threshold for statistical significance.

other variations can be performed:
test if two samples equal ; unequal variances
have the same mean.

Calculate in class from a LL ratio?

Gaussian Approximation to Poisson

$$P(K, \lambda) = \frac{e^{-\lambda}\lambda^K}{K!}$$

mean $= \lambda$

variance $= \lambda$

Approximate using Gaussian distribution with

$\sigma^2 \to \lambda$

$\mu \to \lambda$

$$L(\theta|\vec{y}) \cong \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-(\mu_n - y_n)^2/2\sigma_n^2}$$

$\sigma_n^2 \to \mu_n$

the "MLE" is $\arg_\theta \max \left[\sum_{n=1}^{N} \left[\frac{-(\mu_n - y_n)^2}{\mu_n} - \frac{1}{2}\ln(2\pi\mu_n^2) \right] \right]$

often the log term is dropped

giving $\chi^2 = \sum_{n=1}^{N} \frac{-(\mu_n - y_n)^2}{\mu_n}$

where χ^2 is minimized to find θ.

This is a 'chi-squared' fitting approach.