The Method of Least Squares (LSQ):

The Method of Least Squares is the most frequently used technique for the estimation of parameters. It goes back to Legendre (1805) and Gauss (1809).

Unlike the M.L.M., the LSQ method has ***NO*** *GENERAL OPTIMUM PROPERTIES* (*e.g.* ***minimum variance***, lack of ***bias***, ...) *EVEN AS* $N \to \infty$. However, for the ***restricted*** class of problems where the dependence on the parameters is ***linear***, the LSQ method produces ***unbiased estimators*** of ***minimum variance***, ***even for small samples***.

Our discussion(s) of the M.L.M. were phrased in terms of "N events" whose measurements were $\underline{x} \equiv (x_1, x_2, \ldots, x_N)$. There, we did not worry about the ***uncertainties*** associated with the individual measurements since, at least for large N, the uncertainties are a feature of the observed / measured random variables $x_1, x_2, \ldots, x_N$ which are ***automatically*** included by the M.L.M.

Here, we will consider the case of P ***independent measurements*** of the ***random variable*** $y(x)$ at the ***points*** $x_1, x_2, \ldots, x_P$. The measurements yield results "$(y_1 \pm \sigma_{y_1})$", "$(y_2 \pm \sigma_{y_2})$", ..., "$(y_P \pm \sigma_{y_P})$", where by these we mean that the experimental value is $y_i(x_i)$ with associated standard deviation $\sigma_{y_i}(x_i)$ at the point x_i. One such type of experiment is a ***counting*** experiment, where we measure: n_1 counts at x_1, n_2 counts at x_2, ... and n_P counts at x_P.

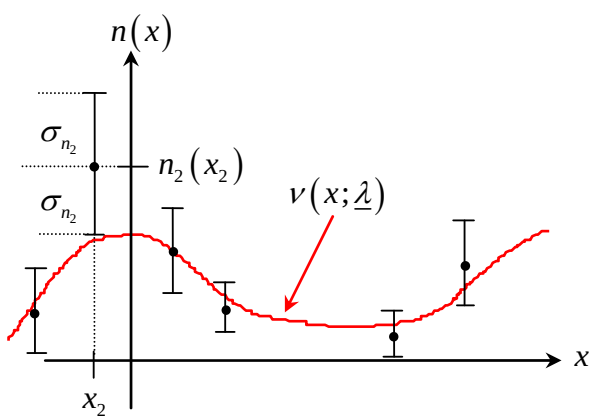
First, we use the M.L.M. We will distinguish between ***two*** cases:

Case I.) $n_i(x_i)$ is ***small*** and Poisson statistics apply:

Here, we assume that the number of measurements P is sufficiently large that the number of counts n_i measured at any particular x_i is small. Then, we ***estimate*** the ***average/true mean*** number of counts expected at x_i as the number n_i of counts actually observed in the measurement. From Poisson statistics, this tells us that: $\sigma_{n_i} = \sqrt{n_i}$.

Recall that: $P(n;\nu) = \dfrac{\nu^n}{n!}e^{-\nu}$ and that: $\hat{n} \equiv E[n] = \sum_{n=0}^{\infty} n \dfrac{\nu^n}{n!} e^{-\nu} = \nu$, $\sigma_n = \sqrt{\nu}$

Suppose we plot the measurements, $n_i(x_i)$ counts *vs.* x_i, *e.g.* associated with a photon-counting double-slit interference experiment, and/or associated with an arbitrary diffraction pattern:



Suppose we have a ***theory*** ***<u>prediction</u>*** for the **red curve**, $\nu\left(x;\underline{\lambda}\right)$ which is the ***average/true mean*** number of counts ***expected*** *vs.* x. Thus, the **red** theory curve tells us the ***average/true mean*** number of counts $\nu_i\left(x_i;\underline{\lambda}\right)$ ***expected*** at ***<u>each</u>*** x_i. Note that there may also be one or more parameters $\underline{\lambda}=\left(\lambda_1,\lambda_2,\ldots\lambda_N\right)$.

If we believe that the theory accurately describes the data, then we want to ***<u>estimate</u>*** the "best" values of the set of parameters $\underline{\lambda}$. We form the ***likelihood*** $\mathcal{L}\left(\underline{\lambda}\right)$ for this problem as follows:

Let $P\left(n;\nu\right)=$ the Poisson probability that we will observe n counts when ν is the average/mean number of expected counts. Then: $P\left(n;\nu\right)=\left(\nu^n/n!\right)e^{-\nu}$.

The probability that n_i counts will occur at x_i is then: $P\left(n_i\left(x_i\right);\nu\left(x_i;\underline{\lambda}\right)\right)=\dfrac{\nu^{n_i\left(x_i\right)}\left(x_i;\underline{\lambda}\right)}{n_i\left(x_i\right)!}e^{-\nu\left(x_i;\underline{\lambda}\right)}$

Thus, for ***<u>all</u>*** of the observed counts, the likelihood $\mathcal{L}_P\left(\underline{\lambda}\right)$ is:

$$\mathcal{L}_P\left(\underline{\lambda}\right)\equiv\prod_{i=1}^{P}P\left(n_i\left(x_i\right);\nu\left(x_i;\underline{\lambda}\right)\right)=\prod_{i=1}^{P}\frac{\nu^{n_i\left(x_i\right)}\left(x_i;\underline{\lambda}\right)}{n_i\left(x_i\right)!}e^{-\nu\left(x_i;\underline{\lambda}\right)}$$

Hence: $\ell_P\left(\underline{\lambda}\right)\equiv\ln\mathcal{L}_P\left(\underline{\lambda}\right)=-\sum_{i=1}^{P}\ln n_i!+\sum_{i=1}^{P}n_i\ln\nu\left(x_i;\underline{\lambda}\right)-\sum_{i=1}^{P}\nu\left(x_i;\underline{\lambda}\right)$

Then: $$\frac{\partial\ell_P\left(\underline{\lambda}\right)}{\partial\lambda_k}=\sum_{i=1}^{P}n_i\frac{1}{\nu\left(x_i;\underline{\lambda}\right)}\frac{\partial}{\partial\lambda_k}\nu\left(x_i;\underline{\lambda}\right)-\sum_{i=1}^{P}\frac{\partial}{\partial\lambda_k}\nu\left(x_i;\underline{\lambda}\right)$$

If we set $\dfrac{\partial\ell_P\left(\underline{\lambda}\right)}{\partial\lambda_k}=0$ for each of the λ_k and then solve each equation to obtain the ***<u>estimates</u>*** $\underline{\lambda}^*=\left(\lambda_1^*,\lambda_2^*,\ldots\lambda_N^*\right)$, we will thus obtain the M.L.M. solution for "Curve Fitting" when there is a ***<u>small</u>*** number of Poisson-distributed counts at each *x*-location.

<u>Case II.)</u> $n_i\left(x_i\right)$ <u>is ***large*** and Gaussian statistics apply</u>:

Here, we assume that the number of counts in each measurement $n_i\left(x_i\right)$ is large, with standard deviation $\sigma_{n_i}\left(x_i\right)$. We assume that the "true" number of counts expected at x_i is Gaussian/normally distributed with P.D.F.: $\dfrac{1}{\sqrt{2\pi}\,\sigma_{n_i}\left(x_i\right)}e^{-\left(n_i\left(x_i\right)-\nu\left(x_i;\underline{\lambda}\right)\right)^2/2\sigma_{n_i}^2\left(x_i\right)}$

Then: $$\mathcal{L}_G\left(\underline{\lambda}\right)\equiv\prod_{i=1}^{P}\frac{1}{\sqrt{2\pi}\,\sigma_{n_i}\left(x_i\right)}e^{-\left[n_i\left(x_i\right)-\nu\left(x_i;\underline{\lambda}\right)\right]^2/2\sigma_{n_i}^2\left(x_i\right)}$$

and: $$\ell_G\left(\underline{\lambda}\right)\equiv\ln\mathcal{L}_G\left(\underline{\lambda}\right)=-\sum_{i=1}^{P}\ln\left(\sqrt{2\pi}\,\sigma_{n_i}\left(x_i\right)\right)-\sum_{i=1}^{P}\frac{\left[n_i\left(x_i\right)-\nu\left(x_i;\underline{\lambda}\right)\right]^2}{2\sigma_{n_i}^2\left(x_i\right)}$$

Thus, in order to ***maximize*** the likelihood $\mathcal{L}_G(\underline{\lambda})$ we must ***minimize***

$$\chi^2(\underline{\lambda}) \equiv \sum_{i=1}^{P} \frac{\left[n_i(x_i) - \nu(x_i;\underline{\lambda})\right]^2}{\sigma_{n_i}{}^2(x_i)} \quad (n.b.\text{ the factor of } \tfrac{1}{2} \text{ is irrelevant here – neglect it!})$$

with respect to the $\underline{\lambda}$-parameters, thereby obtaining the ***estimates*** $\underline{\lambda}^*$. This is the M.L.M. result for "curve fitting" when the number of counts at each *x*-location is Gaussian/normally-distributed.

The Least Squares Principle:

We carry out a total of P ***independent*** measurements – one each at the points $\underline{x} \equiv (x_1, x_2, \ldots, x_P)$ and obtain the experimental values $\underline{y}(\underline{x}) \equiv (y_1(x_1), y_2(x_2), \ldots, y_P(x_P))$, with associated ***apriori known*** standard deviations $\sigma_{\underline{y}}(\underline{x}) \equiv (\sigma_{y_1}(x_1), \sigma_{y_2}(x_2), \ldots, \sigma_{y_P}(x_P))$.

The ***true*** values $\hat{\underline{y}}(\underline{x}) \equiv (\hat{y}_1(x_1), \hat{y}_2(x_2), \ldots, \hat{y}_P(x_P))$ are ***not apriori known***, but *e.g.* a ***theoretical model*** exists which ***predicts*** the "***true***" value $\overline{y}_i(x_i;\underline{\lambda})$ associated with a ***specific*** x_i, where $\underline{\lambda} = (\lambda_1, \lambda_2, \ldots \lambda_N)$ is a set of parameters, ***with the restriction*** that the number of parameters $N \le P$, the number of measurements. Thus, theory predicts the $\overline{\underline{y}}(\underline{x}) \equiv (\overline{y}_1(x_1), \overline{y}_2(x_2), \ldots, \overline{y}_P(x_P))$
Then the "best" values of the $\underline{\lambda}$-parameters are those which ***minimize***

$$\boxed{\chi^2(\underline{\lambda}) \equiv \sum_{i=1}^{P} w_i \left[y_i(x_i) - \overline{y}(x_i;\underline{\lambda})\right]^2}$$ where w_i is the ***weight*** given to the ith measurement.

We see that the LSQ method and the M.L.M. will give the ***same/identical*** $\underline{\lambda}^* = (\lambda_1^*, \lambda_2^*, \ldots \lambda_N^*)$ ***if*** the $y_i(x_i)$ are Gaussian/normally-distributed **.and.** we ***also*** choose $w_i = 1/\sigma_{y_i}^2(x_i)$.

Then:
$$\boxed{\chi^2(\underline{\lambda}) \equiv \sum_{i=1}^{P} \frac{\left[y_i(x_i) - \overline{y}(x_i;\underline{\lambda})\right]^2}{\sigma_{y_i}^2(x_i)}}$$

For Gaussian-distributed measurements, we can easily generalize this result to situations where the P measurements of the $y_i(x_i)$ are ***not*** independent:

$$\boxed{\chi^2(\underline{\lambda}) = \sum_{i=1}^{P}\sum_{j=1}^{P} \left[y_i(x_i) - \overline{y}(x_i;\underline{\lambda})\right] \left(\underline{V}^{-1}_{\underline{y}(\underline{x})}\right)_{ij} \left[y_i(x_i) - \overline{y}(x_i;\underline{\lambda})\right]}$$

where $\underline{V}^{-1}_{\underline{y}(\underline{x})}$ is the ***inverse*** of the ***covariance matrix*** $\underline{V}_{\underline{y}(\underline{x})}$ of the ***measurements*** $y_i(x_i)$.

Formally, recall that since $\overline{y}(x_i;\underline{\lambda}) =$ the "true" value of $y_i(x_i)$ at the point x_i, then the *i-j*th elements of the ***covariance*** matrix of the ***measurements*** $y_i(x_i)$ are:

$$\left(\underline{V}_{\underline{y}(\underline{x})}\right)_{ij} = E\left[\left\{y_i(x_i) - \overline{y}(x_i;\underline{\lambda})\right\}\left\{y_j(x_j) - \overline{y}(x_j;\underline{\lambda})\right\}\right]$$

We can write the P measured $y_i(x_i)$ and P associated theory predictions $\bar{y}(x_i;\underline{\lambda})$ as $P\times 1$ column vectors (*i.e.* $P\times 1$ column matrices):

$$\text{Measured: } \underline{y}(\underline{x}) \equiv \begin{pmatrix} y_1(x_1) \\ y_2(x_2) \\ \vdots \\ y_P(x_P) \end{pmatrix} \qquad \text{Theory Prediction: } \underline{\bar{y}}(\underline{x};\underline{\lambda}) \equiv \begin{pmatrix} \bar{y}(x_1;\underline{\lambda}) \\ \bar{y}(x_2;\underline{\lambda}) \\ \vdots \\ \bar{y}(x_P;\underline{\lambda}) \end{pmatrix}$$

We can then define the ***residual*** $P\times 1$ column matrix as:

$$\underline{R}_{\underline{\lambda}}(\underline{x};\underline{\lambda}) \equiv \underline{y}(\underline{x}) - \underline{\bar{y}}(\underline{x};\underline{\lambda}) \equiv \begin{pmatrix} y_1(x_1) \\ y_2(x_2) \\ \vdots \\ y_P(x_P) \end{pmatrix} - \begin{pmatrix} \bar{y}(x_1;\underline{\lambda}) \\ \bar{y}(x_2;\underline{\lambda}) \\ \vdots \\ \bar{y}(x_P;\underline{\lambda}) \end{pmatrix} = \begin{pmatrix} y_1(x_1) - \bar{y}(x_1;\underline{\lambda}) \\ y_2(x_2) - \bar{y}(x_2;\underline{\lambda}) \\ \vdots \\ y_P(x_P) - \bar{y}(x_P;\underline{\lambda}) \end{pmatrix}$$

and its $1\times P$ row matrix ***<u>transpose</u>*** as:

$$\underline{R}^T_{\underline{\lambda}}(\underline{x};\underline{\lambda}) \equiv \underline{y}^T(\underline{x}) - \underline{\bar{y}}^T(\underline{x};\underline{\lambda}) \equiv \begin{pmatrix} y_1(x_1) \\ y_2(x_2) \\ \vdots \\ y_P(x_P) \end{pmatrix}^T - \begin{pmatrix} \bar{y}(x_1;\underline{\lambda}) \\ \bar{y}(x_2;\underline{\lambda}) \\ \vdots \\ \bar{y}(x_P;\underline{\lambda}) \end{pmatrix}^T = \begin{pmatrix} y_1(x_1) - \bar{y}(x_1;\underline{\lambda}) \\ y_2(x_2) - \bar{y}(x_2;\underline{\lambda}) \\ \vdots \\ y_P(x_P) - \bar{y}(x_P;\underline{\lambda}) \end{pmatrix}^T$$

Then we can write $\chi^2(\underline{\lambda}) = \sum_{i=1}^{P}\sum_{j=1}^{P}\left[y_i(x_i) - \bar{y}(x_i;\underline{\lambda})\right]\left(\underline{V}^{-1}_{\underline{y}(\underline{x})}\right)_{ij}\left[y_i(x_i) - \bar{y}(x_i;\underline{\lambda})\right]$ in matrix notation as:

$$\chi^2(\underline{\lambda}) = \left[\underline{y}(\underline{x}) - \underline{\bar{y}}(\underline{x};\underline{\lambda})\right]^T \left(\underline{V}^{-1}_{\underline{y}(\underline{x})}\right)_{ij}\left[\underline{y}(\underline{x}) - \underline{\bar{y}}(\underline{x};\underline{\lambda})\right] = \underline{R}^T_{\underline{\lambda}}(\underline{x};\underline{\lambda})\underline{V}^{-1}_{\underline{y}(\underline{x})}\underline{R}_{\underline{\lambda}}(\underline{x};\underline{\lambda})$$

<u>Another Perspective of the Least Squares Method</u>:

Suppose we ***measure*** the $\underline{y}(\underline{x}) \equiv \left(y_1(x_1), y_2(x_2), \ldots, y_P(x_P)\right)$ and we wish to know / determine their "***<u>true</u>***" values $\underline{\bar{y}}(\underline{x}) \equiv \left(\bar{y}_1(x_1), \bar{y}_2(x_2), \ldots, \bar{y}_P(x_P)\right)$.

For each ***measurement***, we write: $y_i(x_i) = \bar{y}_i(x_i) + \varepsilon_i(x_i)$, where $\varepsilon_i(x_i)$ is the "error" (*aka* the "***residual***") in the measurement of y_i at the point x_i, *i.e.* $\varepsilon_i(x_i) = y_i(x_i) - \bar{y}_i(x_i)$.

In the absence of any information about the ***<u>specific</u>*** "error" on any measurement (*n.b.* $\varepsilon_i(x_i)$ is ***<u>not</u> apriori known***), we assume that we can still ***<u>estimate</u>*** a standard deviation $\sigma_{y_i}(x_i)$ for that measurement.

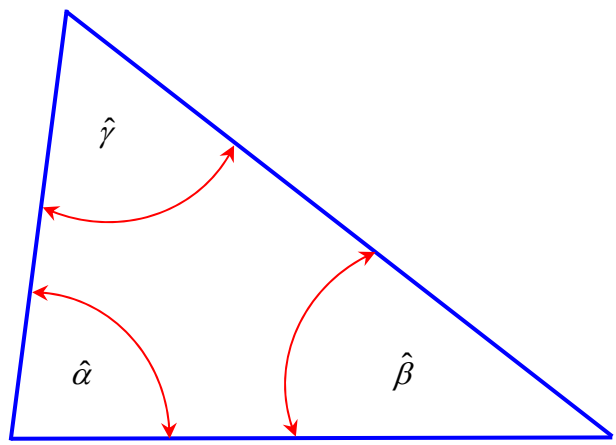
Then we form: $$\chi^2 \equiv \sum_{i=1}^{P}\left(\frac{\varepsilon_i(x_i)}{\sigma_{y_i}(x_i)}\right)^2 = \sum_{i=1}^{P}\frac{\left[y_i(x_i) - \bar{y}_i(x_i)\right]^2}{\sigma^2_{y_i}(x_i)}.$$

and then we determine the "best" ***<u>estimates</u>*** of the $\varepsilon_i(x_i)$ as those that ***minimize*** the above χ^2.

If the theory predictions $\underline{\bar{y}}(\underline{x}) \equiv (\bar{y}_1(x_1), \bar{y}_2(x_2), \ldots, \bar{y}_P(x_P))$ are ***independent***, the solution is trivial: simply choose $\varepsilon_i(x_i) = y_i(x_i) - \bar{y}_i(x_i) = 0$, *i.e.* we choose $y_i(x_i)$ as the ***estimator*** of $\bar{y}_i(x_i)$. (This certainly ***minimizes*** χ^2, which ***must*** be ≥ 0.)

However, consider what happens if the theory predictions $\underline{\bar{y}}(\underline{x}) \equiv (\bar{y}_1(x_1), \bar{y}_2(x_2), \ldots, \bar{y}_P(x_P))$ are ***not independent***, for example if they are subject to ***constraints***.

The classic example is the famous Surveyor's "***Failure to Close***" problem – *e.g.* measuring the three internal angles of a (plane) triangle. Let $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ be the ***true*** (but ***apriori unknown***) internal angles of a particular triangle, as shown in the figure below:



Suppose our (***independent***!) experimental measurements of the three angles are: $\alpha \pm \sigma_\alpha$, $\beta \pm \sigma_\beta$ and $\gamma \pm \sigma_\gamma$. We ***apriori know*** from trigonometry that $\hat{\alpha} + \hat{\beta} + \hat{\gamma} = \pi$ for the ***true*** angles. But do our angle ***measurements also*** satisfy $\alpha + \beta + \gamma = \pi$? The surveyor's "***failure-to-close***" on the three internal angle ***measurements*** means that $|(\alpha + \beta + \gamma) - \pi| > 0$, *i.e.* there is either an ***excess*** or a ***deficit*** "***failure-to-close***" angle ***residual***: $\Delta \equiv \pi - (\alpha + \beta + \gamma)$.

We want to obtain ***estimates*** for the "true" angles $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$. How do we go about doing this?

Let us begin by first ***ignoring*** the ***angle constraint*** $\bar{\alpha} + \bar{\beta} + \bar{\gamma} = \pi$. The χ^2 for this situation is:

$$\chi^2(\bar{\alpha}, \bar{\beta}, \bar{\gamma}) = \frac{(\alpha - \bar{\alpha})^2}{\sigma_\alpha^2} + \frac{(\beta - \bar{\beta})^2}{\sigma_\beta^2} + \frac{(\gamma - \bar{\gamma})^2}{\sigma_\gamma^2}$$

We then ***minimize*** this χ^2 expression by taking derivatives *w.r.t.* the "true" angle parameters:

$$\frac{\partial \chi^2}{\partial \bar{\alpha}} = -2\frac{(\alpha - \bar{\alpha})}{\sigma_\alpha^2} = 0, \quad \frac{\partial \chi^2}{\partial \bar{\beta}} = -2\frac{(\beta - \bar{\beta})}{\sigma_\beta^2} = 0 \text{ and: } \frac{\partial \chi^2}{\partial \bar{\gamma}} = -2\frac{(\gamma - \bar{\gamma})}{\sigma_\gamma^2} = 0$$

Which gives: $\bar{\alpha} = \alpha$, $\bar{\beta} = \beta$ and: $\bar{\gamma} = \gamma$. ***Clearly, this is wrong!***

If we now ***include*** the ***angle constraint*** in order to ***eliminate*** one of the three parameters, *e.g.* $\bar{\gamma} = \pi - \bar{\alpha} - \bar{\beta}$, then the χ^2 expression becomes:

$$\chi^2\left(\bar{\alpha},\bar{\beta}\right) = \frac{\left(\alpha-\bar{\alpha}\right)^2}{\sigma_\alpha^2} + \frac{\left(\beta-\bar{\beta}\right)^2}{\sigma_\beta^2} + \frac{\left(\gamma-\left(\pi-\bar{\alpha}-\bar{\beta}\right)\right)^2}{\sigma_\gamma^2}$$

We ***minimize*** ***this*** χ^2 expression by taking derivatives *w.r.t.* the two remaining "***true***" angle parameters $\bar{\alpha}$ and $\bar{\beta}$:

$$\frac{\partial\chi^2}{\partial\bar{\alpha}} = -2\frac{\left(\alpha-\bar{\alpha}\right)}{\sigma_\alpha^2} + 2\frac{\left(\bar{\alpha}+\bar{\beta}+\gamma-\pi\right)}{\sigma_\gamma^2} = 0 \quad \text{or:} \quad \left(\bar{\alpha}-\alpha\right) = \left(\pi-\bar{\alpha}-\bar{\beta}-\gamma\right)\frac{\sigma_\alpha^2}{\sigma_\gamma^2} \qquad [1^{\text{o}}]$$

$$\frac{\partial\chi^2}{\partial\bar{\beta}} = -2\frac{\left(\beta-\bar{\beta}\right)}{\sigma_\beta^2} + 2\frac{\left(\bar{\alpha}+\bar{\beta}+\gamma-\pi\right)}{\sigma_\gamma^2} = 0 \quad \text{or:} \quad \left(\bar{\beta}-\beta\right) = \left(\pi-\bar{\alpha}-\bar{\beta}-\gamma\right)\frac{\sigma_\beta^2}{\sigma_\gamma^2} \qquad [2^{\text{o}}]$$

Next, we need to solve these equations simultaneously… First, we get rid of the denominators on the RHS of the above two equations:

$$\left(\bar{\alpha}-\alpha\right)\sigma_\gamma^2 = \left(\pi-\bar{\alpha}-\bar{\beta}-\gamma\right)\sigma_\alpha^2 \quad [1] \quad \text{and:} \quad \left(\bar{\beta}-\beta\right)\sigma_\gamma^2 = \left(\pi-\bar{\alpha}-\bar{\beta}-\gamma\right)\sigma_\beta^2 \quad [2]$$

But: $\bar{\alpha}+\bar{\beta}+\bar{\gamma} = \pi$, or: $\bar{\gamma} = \pi - \bar{\alpha} - \bar{\beta}$. Then these two equations become:

$$\left(\bar{\alpha}-\alpha\right)\sigma_\gamma^2 = \left(\bar{\gamma}-\gamma\right)\sigma_\alpha^2 \quad [1'] \quad \text{and:} \quad \left(\bar{\beta}-\beta\right)\sigma_\gamma^2 = \left(\bar{\gamma}-\gamma\right)\sigma_\beta^2 \quad [2']$$

Now take the ***ratio*** of these two equations to obtain a ***third*** {***similar***} relation:

$$\frac{\left(\bar{\alpha}-\alpha\right)\cancel{\sigma_\gamma^2}}{\left(\bar{\beta}-\beta\right)\cancel{\sigma_\gamma^2}} = \frac{\cancel{\left(\bar{\gamma}-\gamma\right)}\sigma_\alpha^2}{\cancel{\left(\bar{\gamma}-\gamma\right)}\sigma_\beta^2} \quad \Rightarrow \quad \left(\bar{\alpha}-\alpha\right)\sigma_\beta^2 = \left(\bar{\beta}-\beta\right)\sigma_\alpha^2 \quad [3']$$

Now ***add***: $\left(\bar{\alpha}-\alpha\right)\left(\sigma_\alpha^2+\sigma_\beta^2\right)$ to ***both*** sides of equation [1]:

$$\left(\bar{\alpha}-\alpha\right)\left(\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2\right) = \left(\pi-\bar{\alpha}-\bar{\beta}-\gamma\right)\sigma_\alpha^2 + \left(\bar{\alpha}-\alpha\right)\left(\sigma_\alpha^2+\sigma_\beta^2\right) \quad [1'']$$

Use equation [3′] to replace $\left(\bar{\alpha}-\alpha\right)\sigma_\beta^2$ with $\left(\bar{\beta}-\beta\right)\sigma_\alpha^2$ in the RHS of equation [1″]:

$$\left(\bar{\alpha}-\alpha\right)\left(\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2\right) = \left[\left(\pi-\cancel{\bar{\alpha}}-\cancel{\bar{\beta}}-\gamma\right)+\left(\cancel{\bar{\alpha}}-\alpha\right)+\left(\cancel{\bar{\beta}}-\beta\right)\right]\sigma_\alpha^2 \quad [1''']$$

$$= \underbrace{\left[\pi-\left(\alpha+\beta+\gamma\right)\right]}_{\equiv\Delta}\sigma_\alpha^2 = \Delta\sigma_\alpha^2$$

Then we see that: $\bar{\alpha} = \alpha + \Delta\left[\dfrac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\right] \quad [1^{\text{iv}}]$

We repeat this ***same*** methodology for equation [2], we obtain: $\bar{\beta} = \beta + \Delta\left[\dfrac{\sigma_\beta^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\right] \quad [2^{\text{iv}}]$

Then:

$$\bar{\gamma} = \pi - \bar{\alpha} - \bar{\beta} = (\pi - \alpha - \beta) - \Delta\left[\frac{\sigma_\alpha^2 + \sigma_\beta^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}\right] = \underbrace{(\pi - \alpha - \beta - \Delta)}_{=\gamma} + \Delta\left[\frac{\sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}\right].$$

$$= \gamma + \Delta\left[\frac{\sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}\right] \quad [3^{\text{iv}}]$$

Thus our LSQ fit ***estimators*** for the three internal angles of the triangle are:

$$\bar{\alpha} = \alpha + \Delta\left[\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}\right]$$

$$\bar{\beta} = \beta + \Delta\left[\frac{\sigma_\beta^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}\right]$$

and:

$$\bar{\gamma} = \gamma + \Delta\left[\frac{\sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2}\right]$$

Thus, we see that the "***failure-to-close***" angle ***residual*** Δ is split ***symmetrically*** between ***each*** of the three internal angle measurements α, β and γ, weighted by the fractional variance of the respective angles in terms of applying small corrections to the angles in order to obtain ***estimates*** of the "***true***" internal angles.

Another method that enables us to explicitly include the angle ***constraint*** $\bar{\alpha} + \bar{\beta} + \bar{\gamma} = \pi$ in the χ^2 ***minimization*** process is to ***add*** a so-called ***Lagrange Multiplier*** term $\lambda \cdot (\bar{\alpha} + \bar{\beta} + \bar{\gamma} - \pi)$ to the χ^2 which then becomes:

$$\chi'^2(\bar{\alpha}, \bar{\beta}, \bar{\gamma}) = \chi^2(\bar{\alpha}, \bar{\beta}, \bar{\gamma}) + \lambda \cdot (\bar{\alpha} + \bar{\beta} + \bar{\gamma} - \pi) = \frac{(\alpha - \bar{\alpha})^2}{\sigma_\alpha^2} + \frac{(\beta - \bar{\beta})^2}{\sigma_\beta^2} + \frac{(\gamma - \bar{\gamma})^2}{\sigma_\gamma^2} + \lambda \cdot (\bar{\alpha} + \bar{\beta} + \bar{\gamma} - \pi)$$

We ***minimize this*** expression by taking derivatives *w.r.t.* the three "true" angle parameters $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ and then setting these expressions = 0:

$$\frac{\partial \chi^2}{\partial \bar{\alpha}} = -2\frac{(\alpha - \bar{\alpha})}{\sigma_\alpha^2} + \lambda = 0, \quad \frac{\partial \chi^2}{\partial \bar{\beta}} = -2\frac{(\beta - \bar{\beta})}{\sigma_\beta^2} + \lambda = 0 \quad \text{and:} \quad \frac{\partial \chi^2}{\partial \bar{\gamma}} = -2\frac{(\gamma - \bar{\gamma})}{\sigma_\gamma^2} + \lambda = 0$$

and:

$$\frac{\partial \chi^2}{\partial \lambda} = (\bar{\alpha} + \bar{\beta} + \bar{\gamma} - \pi) = 0$$

Then: $\bar{\alpha} = \alpha + \frac{1}{2}\lambda\sigma_\alpha^2$, $\quad \bar{\beta} = \beta + \frac{1}{2}\lambda\sigma_\beta^2 \quad$ and: $\quad \bar{\gamma} = \gamma + \frac{1}{2}\lambda\sigma_\gamma^2$

Next, we add these last three equations together: $(\bar{\alpha} + \bar{\beta} + \bar{\gamma}) = (\alpha + \beta + \gamma) + \frac{1}{2}\lambda \cdot (\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2)$.

But: $\left(\bar{\alpha}+\bar{\beta}+\bar{\gamma}\right)=\pi$ from the ***Lagrange Multiplier constraint***, and also: $\Delta \equiv \pi-\left(\alpha+\beta+\gamma\right)$.

Thus: $\pi-\left(\alpha+\beta+\gamma\right)=\Delta=\frac{1}{2}\lambda\cdot\left(\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2\right)$ and hence the ***Lagrange Multiplier*** is:

$$\lambda=\frac{2\Delta}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}.$$

Hence, we obtain the LSQ + ***Lagrange Multiplier*** fit results for the ***<u>estimates</u>*** of the "***<u>true</u>***" values of the three internal angles of the triangle:

$$\bar{\alpha}=\alpha+\tfrac{1}{2}\lambda\sigma_\alpha^2=\alpha+\Delta\left[\frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\right]$$

$$\bar{\beta}=\beta+\tfrac{1}{2}\lambda\sigma_\beta^2=\beta+\Delta\left[\frac{\sigma_\beta^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\right]$$

$$\bar{\gamma}=\gamma+\tfrac{1}{2}\lambda\sigma_\gamma^2=\gamma+\Delta\left[\frac{\sigma_\gamma^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\right]$$

We (again) see that the ***excess*** or ***deficit*** "***failure-to-close***" angle ***<u>residual</u>***: $\Delta \equiv \pi-\left(\alpha+\beta+\gamma\right)$ is shared symmetrically among all of the measurements, with the "best" measurement (*i.e.* the one with the ***smallest*** angle uncertainty σ) being changed the ***least***.

These expressions are identical to those obtained above, by directly including the angle ***constraint*** in the (two parameter) LSQ fit χ^2 expression, as we expect.

This problem is an example of the "***Adjustment of Parameters***" by the LSQ + ***Lagrange Multiplier*** method, where ***<u>constraints</u>*** among the ***<u>measured</u>*** quantities are used to ***reduce*** the ***uncertainties*** on them.

The ***original/initial*** 3×3 ***covariance*** matrix of ***parameters*** for this problem is ***<u>diagonal</u>***, since the internal angle ***measurements*** α, β and γ are ***<u>independent</u>*** ***random variables***:

$$V_{orig}=\begin{pmatrix}\sigma_\alpha^2 & 0 & 0\\ 0 & \sigma_\beta^2 & 0\\ 0 & 0 & \sigma_\gamma^2\end{pmatrix}$$

Thus, the ***<u>inverse</u>*** of the ***original/initial*** 3×3 ***covariance*** matrix of ***parameters*** is:

$$V_{orig}^{-1}=\begin{pmatrix}1/\sigma_\alpha^2 & 0 & 0\\ 0 & 1/\sigma_\beta^2 & 0\\ 0 & 0 & 1/\sigma_\gamma^2\end{pmatrix}$$

However, ***after*** the LSQ fit is carried out, the "new" ***covariance*** matrix of ***parameters*** (which can be obtained from error propagation) is ***non-diagonal*** (*cf*. P598AEM Lect. Notes 20, *p*. 14-15):

$$\hat{V}_{fit} = V_{orig} - \frac{1}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\begin{pmatrix}\sigma_\alpha^2\\ \sigma_\beta^2\\ \sigma_\gamma^2\end{pmatrix}\begin{pmatrix}\sigma_\alpha^2 & \sigma_\beta^2 & \sigma_\gamma^2\end{pmatrix}$$

$$= \begin{pmatrix}\sigma_\alpha^2 & 0 & 0\\ 0 & \sigma_\beta^2 & 0\\ 0 & 0 & \sigma_\gamma^2\end{pmatrix} - \frac{1}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\begin{pmatrix}\sigma_\alpha^4 & \sigma_\alpha^2\sigma_\beta^2 & \sigma_\alpha^2\sigma_\gamma^2\\ \sigma_\alpha^2\sigma_\beta^2 & \sigma_\beta^4 & \sigma_\beta^2\sigma_\gamma^2\\ \sigma_\alpha^2\sigma_\gamma^2 & \sigma_\beta^2\sigma_\gamma^2 & \sigma_\gamma^4\end{pmatrix}$$

$$= \begin{pmatrix}\sigma_\alpha^2\left(1-\frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\right) & -\frac{\sigma_\alpha^2\sigma_\beta^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2} & -\frac{\sigma_\alpha^2\sigma_\gamma^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\\ -\frac{\sigma_\alpha^2\sigma_\beta^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2} & \sigma_\beta^2\left(1-\frac{\sigma_\beta^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\right) & -\frac{\sigma_\beta^2\sigma_\gamma^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\\ -\frac{\sigma_\alpha^2\sigma_\gamma^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2} & -\frac{\sigma_\beta^2\sigma_\gamma^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2} & \sigma_\gamma^2\left(1-\frac{\sigma_\gamma^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2}\right)\end{pmatrix}$$

Note that the diagonal elements of the "***new***" ***covariance*** matrix of ***parameters*** $\hat{V}_{fit}$ are ***less*** (*i.e*. ***smaller***) than those of the ***original*** covariance matrix V_{orig}, but the "***new***" ***covariance*** matrix of ***parameters*** now also has non-zero, negative off-diagonal elements, due to the presence of (anti-) correlations of the fitted parameters – *i.e*. increasing (decreasing) one angle causes a decrease (increase) in the other two angles, respectively.

Now for a more general treatment of the LSQ method. We distinguish three cases:

a.) The $y_i(x_i)$ ***measurements*** are ***Gaussian*** distributed.

b.) The "theory" ***prediction*** function $\overline{y}(x_i;\underline{\lambda})$ is ***linear*** in the $\underline{\lambda}$-parameters.

c.) The "theory" ***prediction*** function $\overline{y}(x_i;\underline{\lambda})$ is ***not*** a ***linear*** function of the $\underline{\lambda}$-parameters.

Case *a*.) The measurements are ***Gaussian*** distributed.

We have seen that ***minimization*** of the LSQ sum $\chi^2(\underline{\lambda})$ is algebraically the same/identical as the ***maximization*** of the likelihood $\mathcal{L}(\underline{\lambda})$. Therefore, we know that the LSQ method produces ***estimators*** $\underline{\lambda}^*$, which are ***minimum variance*** and ***unbiased***. (Recall that for Gaussian/normally distributed measurements the M.L.M. has this property.) But there some ***new*** things to be learned about the nature of the $\chi^2(\underline{\lambda})$ function.

Let's assume a that set of P ***independent*** quantities $\underline{\overline{y}}(\underline{x}) \equiv (\overline{y}_1(x_1), \overline{y}_2(x_2), \ldots, \overline{y}_P(x_P))$ exists that are ***apriori known*** – *e.g.* a theoretical model exists which ***predicts*** the $\overline{y}_i(x_i)$, and that a set of P ***independent measurements*** of these quantities also exists: $\underline{y}(\underline{x}) \equiv (y_1(x_1), y_2(x_2), \ldots, y_P(x_P))$ along with their associated ***apriori known*** standard deviations $\underline{\sigma}_y(x) \equiv (\sigma_{y_1}(x_1), \sigma_{y_2}(x_2), \ldots, \sigma_{y_P}(x_P))$.

We additionally assume that the ***independent measurements*** are Gaussian distributed as $N(\overline{y}, \sigma_y)$.

We can then define a chi-squared statistic $\chi^2(\underline{\lambda})$, which is physically the sum of the squares of the ***fractional residuals*** (*i.e.* normalized by the experimental standard deviations):

$$\chi^2(\underline{\lambda}) \equiv \sum_{i=1}^{P} \frac{\left[y_i(x_i) - \overline{y}(x_i;\underline{\lambda})\right]^2}{\sigma_{y_i}^2(x_i)}$$

Again, if we conceptually imagine repeating the ***entire*** experiment of P independent measurements of $\underline{y}(\underline{x}) \equiv (y_1(x_1), y_2(x_2), \ldots, y_P(x_P))$ a gazillion times, we see that the chi-squared statistic $\chi^2(\underline{\lambda})$ is ***also*** a ***random variable***; *i.e.* it is randomly distributed in accordance with its own Probability Density Function (P.D.F) for P ***degrees of freedom***:

$$f(\chi^2, P) = \frac{1}{2^{\frac{P}{2}}\Gamma\left(\frac{P}{2}\right)} (\chi^2)^{\frac{P}{2}-1} e^{-\chi^2/2}$$

Proof: let $u_i \equiv \dfrac{y_i - \overline{y}_i}{\sigma_{y_i}}$, which is Gaussian/normal distributed as $N(0,1)$.

$\therefore$ u_i has P.D.F. $$f(u_i) = \frac{1}{\sqrt{2\pi}} e^{-u_i^2/2}$$

$\therefore$ the ***joint*** P.D.F. for $\underline{u} \equiv (u_1, u_2, \ldots, u_P)$ is: $f(\underline{u}) = \prod_{i=1}^{P} f(u_i) = \frac{1}{(2\pi)^{\frac{P}{2}}} e^{-\left(u_1^2+u_2^2+\ldots+u_P^2\right)/2} = \frac{1}{(2\pi)^{\frac{P}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{P} u_i^2}$

We now define: $z \equiv u_1^2 + u_2^2 + \ldots + u_2^2 = \sum_{i=1}^{P} u_i^2$, then we can write: $f(\underline{u}) = \frac{1}{(2\pi)^{\frac{P}{2}}} e^{-z/2}$

We then change variables to work only in terms of the z-variable, *i.e.* we set:

$$f(\underline{u})\,d\underline{u} \equiv f(u_1, u_2, \ldots, u_P)\,du_1 du_2 \ldots du_P = g(z)\,dz$$

In the P-dimensional "$u_1, u_2, \ldots, u_P$" space we have: $z \equiv u_1^2 + u_2^2 + \ldots + u_2^2 = \sum_{i=1}^{P} u_i^2 \equiv \rho^2$,

where ρ is the ***radius*** of a ***P-dimensional sphere*** and $dV \equiv du_1 du_2 \ldots du_P$ is the ***volume element*** of a P-dimensional sphere in ***Cartesian coordinates***.

However, we want/need the ***volume element*** in ***spherical coordinates***…

The ***volume*** V of the P-dimensional sphere ***must*** be proportional to ρ^P, *i.e.* $V = k\rho^P$.

$\therefore$ $dV = kP\rho^{P-1} d\rho$, and note also that since $z = \rho^2$, then: $dz = 2\rho d\rho$, or: $d\rho = \frac{dz}{2\rho} = \frac{dz}{2\sqrt{z}}$

$\therefore$ $dV = kP\rho^{P-1} d\rho = kPz^{\frac{P-1}{2}} \frac{dz}{2\sqrt{z}} = \frac{1}{2} kPz^{\frac{P-1}{2}} z^{-\frac{1}{2}} dz = \frac{1}{2} kPz^{\frac{P}{2}-\frac{1}{2}} z^{-\frac{1}{2}} dz = \frac{1}{2} kPz^{\frac{P}{2}-1} dz$ or: $\frac{dV}{dz} = \frac{1}{2} kPz^{\frac{P}{2}-1}$

Now: $f(\underline{u})\,d\underline{u} \equiv f(u_1, u_2, \ldots, u_P)\,du_1 du_2 \ldots du_P = f(\underline{u})\,dV = g(z)\,dz$

Then: $g(z) = f(\underline{u}) \frac{dV}{dz} = k\left(\frac{P}{2}\right) z^{\frac{P}{2}-1} \frac{1}{(2\pi)^{\frac{P}{2}}} e^{-z/2}$, or: $g(z) = k' z^{\frac{P}{2}-1} e^{-z/2}$

We determine k' by the normalization requirement for P.D.F.'s: $\int_0^\infty g(z)\,dz = 1$ gives: $k' = \frac{1}{2^{\frac{P}{2}} \Gamma\left(\frac{P}{2}\right)}$

The gamma function is defined as: $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$.

Note that for integer n: $\Gamma(n) = (n-1)!$ and: $\Gamma(x+1) = x\Gamma(x)$, and: $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Then:
$$g(z) = \frac{1}{2^{\frac{P}{2}} \Gamma\left(\frac{P}{2}\right)} z^{\frac{P}{2}-1} e^{-z/2}$$

Finally, if we replace the variable z by the chi-squared variable $\chi^2 = z$, then the chi-squared P.D.F. is:

$$f(\chi^2; P) = \frac{1}{2^{\frac{P}{2}} \Gamma\left(\frac{P}{2}\right)} \left(\chi^2\right)^{\frac{P}{2}-1} e^{-\chi^2/2}$$

<u>Some Properties of the χ^2 Probability Distribution Function (P.D.F)</u>:

$$f\left(\chi^2;P\right)=\frac{1}{2^{\frac{P}{2}}\Gamma\left(\frac{P}{2}\right)}\left(\chi^2\right)^{\frac{P}{2}-1}e^{-\chi^2/2}$$

Define: $h\equiv\frac{1}{2}P$ (h stands for "half") and: $\nu\equiv\frac{1}{2}\chi^2$.

Then: $d\left(\chi^2\right)=2d\nu$ and: $f\left(\chi^2;P\right)d\left(\chi^2\right)=\frac{1}{2^h\Gamma\left(h\right)}\nu^{h-1}e^{-\nu}d\nu=g\left(\nu\right)d\nu$

Thus:
$$g\left(\nu\right)=\frac{1}{2^h\Gamma\left(h\right)}\nu^{h-1}e^{-\nu}.$$

Since P is an integer, then h is either an integer or half-integer. For $h\ \geq\ 3/2$ (*i. e.* $P\geq 3$) the function $g\left(\nu\right)$ increases as ν^{h-1} and decreases as $e^{-\nu}$. $\therefore$ $g\left(\nu\right)$ has a ***maximum*** at some $\nu>0$.

For $h = 1$ ($P = 2$), $g\left(\nu\right)=e^{-\nu}$, which has a ***maximum*** at $\nu=0$.
For $h = 1/2$ ($P = 1$), $g\left(\nu\right)=e^{-\nu}\big/\sqrt{\nu}$ which $\rightarrow\infty$ (*i.e.* becomes ***singular***) as $\nu\rightarrow 0$.

Let us determine where the ***maximum*** occurs for $h\geq 3/2$ ($P\geq 3$) by setting $\frac{dg\left(\nu\right)}{d\nu}=0$:

$$\frac{dg\left(\nu\right)}{d\nu}=\left[\frac{\left(h-1\right)}{\nu}-1\right]\frac{1}{2^h\Gamma\left(h\right)}\nu^{h-1}e^{-\nu}=0\ \Rightarrow\ \left[\frac{\left(h-1\right)}{\nu}-1\right]=0\ \Rightarrow\ \nu_{\max}=\left(h-1\right)$$

$\therefore$ The ***maximum*** (*i.e.* the ***mode*** = ***most probable value***) occurs at $\nu_{\max}=\left(h-1\right)$.
Thus, since $h\equiv\frac{1}{2}P$ and $\nu\equiv\frac{1}{2}\chi^2$, the ***maximum*** of the χ^2 P.D.F. $f\left(\chi^2;P\right)$ occurs at $\chi^2_{\max}=2\left(h-1\right)=2h-2=P-2$ if $P\geq 2$.

n.b. For $P = 2$, $\chi^2_{\max}=0$ is not a "local maximum", $\therefore$ For $P = 2$, the ***most probable value*** of $\chi^2=0$.

Let's calculate the ***expectation value*** (*i.e.* ***mean/average value***) of χ^2:

$$E[\nu]=\int_0^\infty \nu g\left(\nu\right)d\nu=\frac{1}{\Gamma(h)}\int_0^\infty \nu^h e^{-\nu}d\nu=\frac{\Gamma(h+1)}{\Gamma(h)}=h=\tfrac{1}{2}P\quad\therefore\ \text{Since } \nu\equiv\tfrac{1}{2}\chi^2\text{, then: } E[\chi^2]=P.$$

Likewise, we can also calculate the ***variance*** of χ^2:

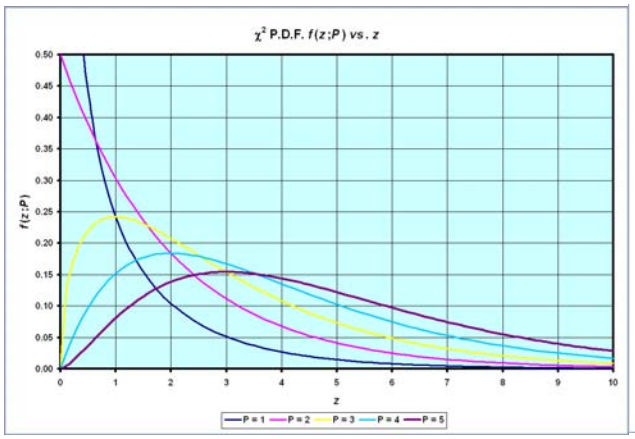
$$E[\nu^2]=\int_0^\infty \nu^2 g\left(\nu\right)d\nu=\frac{1}{\Gamma(h)}\int_0^\infty \nu^{h+1}e^{-\nu}d\nu=\frac{\Gamma(h+2)}{\Gamma(h)}=h\left(h+1\right)=\tfrac{1}{2}P\left(\tfrac{1}{2}P+1\right)$$

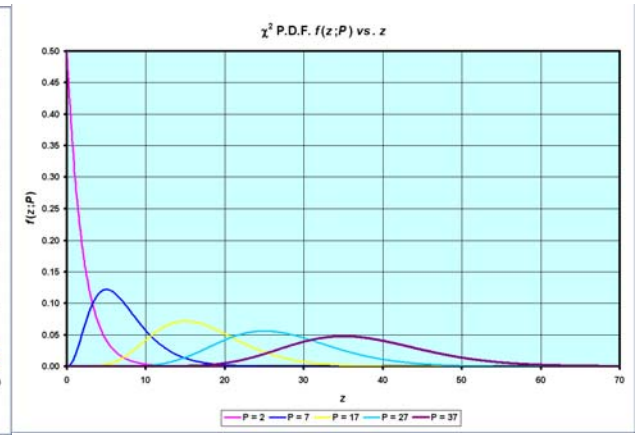
$$\therefore\qquad E[\left(\chi^2\right)^2]=E[\left(2\nu\right)^2]=4h\left(h+1\right)=2P\left(\tfrac{1}{2}P+1\right)=P\left(P+2\right)$$

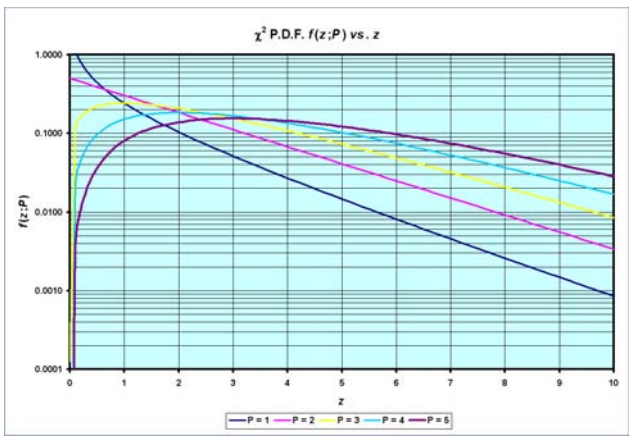
$$\therefore\qquad \sigma^2_{\chi^2}=E[\left(\chi^2\right)^2]-\left(E[\chi^2]\right)^2=P\left(P+2\right)-P^2=2P$$

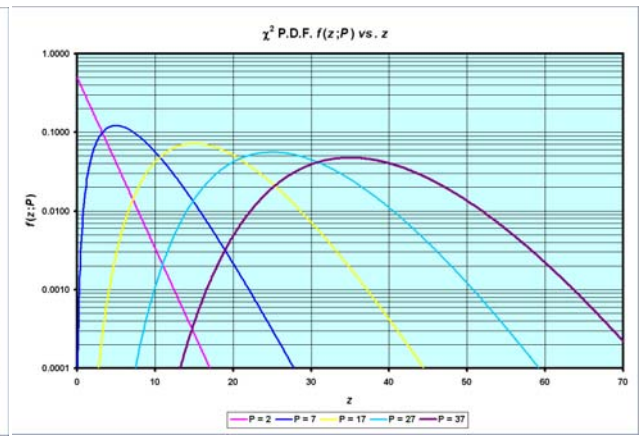
Higher moments of the χ^2 P.D.F. can also be calculated – for reference, the ***skewness*** of the χ^2 distribution is $\sqrt{8/P}$ and the ***excess kurtosis*** of the χ^2 distribution is 12/*P*.

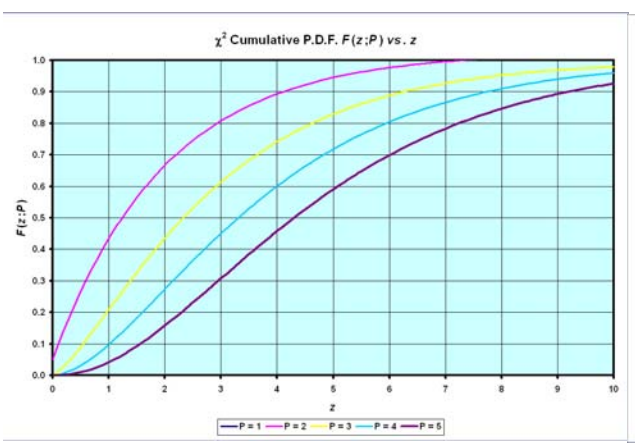
The figures below show plots of various χ^2 P.D.F.'s and their ***cumulative*** P.D.F's (bottom):

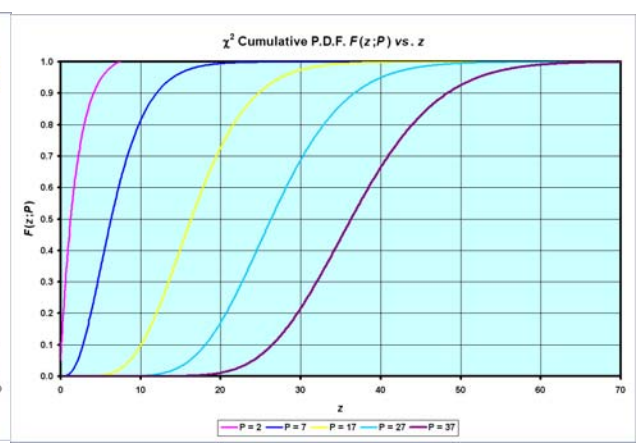












If we look at the above plots of $f\left(\chi^2;P\right)$, we see that the expected properties are apparent: For $P>2$ the ***maximum*** (*i.e.* the ***mode*** – the ***most probable value***) occurs at $z=\chi^2=P-2$; the ***mean*** occurs at $E[z]=E[\chi^2]=P$ and the width grows with P as: $\sigma_{\chi^2}=\sqrt{\sigma^2_{\chi^2}}=\sqrt{2P}$. As P increases, the χ^2 distribution(s) become more and more Gaussian.

The ***characteristic function*** associated with the χ^2 probability distribution function is:

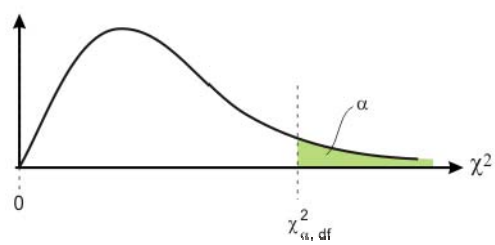
$$\Phi_P\left(k\right)\equiv\int_0^\infty e^{ikz}f\left(z;P\right)dz=\int_0^\infty e^{ikz}\frac{1}{2^{\frac{P}{2}}\Gamma\left(\frac{P}{2}\right)}z^{\frac{P}{2}-1}e^{-z/2}dz=\left(1-2ik\right)^{-\frac{P}{2}}$$

Using the χ^2 ***characteristic function***, it is straightforward to show that if P is ***large***, the ***random variable*** $\alpha_1\equiv\left(\chi^2-\left\langle\chi^2\right\rangle\right)\Big/\sigma_{\chi^2}=\left(\chi^2-P\right)\Big/\sqrt{2P}$ ***approximately*** follows the unit Normal $N(0,1)$ distribution. Quantitatively: $f\left(\chi^2;P\right)d\chi^2=\frac{1}{\sqrt{2\pi}\sqrt{2P}}e^{-\left(\chi^2-P\right)^2/4P}d\chi^2$ for $P\geq 30$.

The agreement improves as $P\to\infty$. The ***random variable*** $\alpha_2\equiv\sqrt{2\chi^2}-\sqrt{2P-1}$ is even closer to $N(0,1)$ in its distribution!

The so-called ***p-value***, *aka* ***Single-Sided Upper Confidence Level*** (CL^{SS}_{upper}) associated with a given/specific χ^2-value and with its number of degrees of freedom, $N_{DoF}=P$ is the integral of the high-side tail of the χ^2 P.D.F. $f\left(z;N_{DoF}\right)$ (see figure below), expressed as a per cent (%):

$$p-value=CL^{SS}_{upper}\equiv 100\times\int_z^\infty f\left(z';N_{DoF}\right)dz'.$$



However, since the ***cumulative*** χ^2 P.D.F. is: $F\left(z;N_{DoF}\right)\equiv\int_0^z f\left(z';N_{DoF}\right)dz'$, then we can equivalently express the $p-value/CL^{SS}_{upper}$ as:

$$p-value=CL^{SS}_{upper}\equiv 100\times\int_z^\infty f\left(z';N_{DoF}\right)dz'=100\times\left[1-F\left(z;N_{DoF}\right)\right]=100\times\left[1-\int_0^z f\left(z';N_{DoF}\right)dz'\right]$$

A $p-value/CL^{SS}_{upper}$ of ~ 0% corresponds to a ***large*** χ^2-value, which is ***bad*** – this warns the experimentalist that the data doesn't agree with the theory – either there ***is*** a fundamental disagreement, or *e.g.* experimental uncertainties $\underline{\sigma}_y\left(\underline{x}\right)$ have been significantly ***underestimated***.

A $p-value/CL^{SS}_{upper}$ of ~ 100% corresponds to a ***low*** χ^2-value, which ***may*** be interpreted as a ***good*** thing, *however*, low χ^2-values can ***also*** arise *e.g.* from ***over-inflated*** experimental uncertainties, $\underline{\sigma}_y\left(\underline{x}\right)$.

Suppose a physicist carries out a measurement of some physical process, obtaining the experimental results $\underline{y}(\underline{x})$ with accompanying $\underline{\sigma}_y(\underline{x})$, which then can be compared *e.g.* to a ***theory prediction*** of $\underline{\bar{y}}(\underline{x};\underline{\lambda})$, by ***minimizing*** the $\chi^2(\underline{\lambda}) = \sum_{i=1}^{P}\left(y_i(x_i) - \bar{y}_i(x_i;\underline{\lambda})\right)^2 \Big/ \sigma_{y_i}^2$, in order to determine the theory parameter ***estimates*** $\underline{\lambda}^*$. For that ***single*** experiment, the resulting chi-squared is $\chi^2(\underline{\lambda}^*)$. If there are N $\underline{\lambda}$-parameters associated with the theory prediction, then the ***number of degrees of freedom*** associated with this χ^2 function is $N_{DoF} = P - N$.

If an experimentalist has done a careful/diligent job on determining his/her data – i.e. ***both*** the central values $\underline{y}(\underline{x})$ ***and*** their associated 1-standard deviation statistical uncertainties $\underline{\sigma}_y(\underline{x})$ **.and.** there are in fact ***no*** experimental biases/no sources of unknown systematic effects, then if one conceptually imagines repeating the entire experiment from start-to-finish a gazillion times, histogramming the χ^2's resulting from the ensemble of a gazillion repetitions of this experiment, the ***shape*** of this (normalized) χ^2 histogram will be the ***same*** as that for the χ^2 P.D.F. with its associated $N_{DoF} = P - N$. The χ^2 statistic ***is*** a ***random variable*** – and ***is*** distributed as the χ^2 P.D.F for $N_{DoF} = P - N$ ***degrees of freedom***.

Furthermore, if one computes the ***cumulative*** χ^2 P.D.F. $F(z_i;N_{DoF}) \equiv \int_0^{z_i} f(z';N_{DoF})\,dz'$ associated with each individual $z_i = \chi_i^2$ resulting from the ensemble of a gazillion repetitions of the experiment, histogramming $F(z_i;N_{DoF}) \equiv \int_0^{z_i} f(z';N_{DoF})\,dz'$, the ***shape*** of the normalized $F(z_i;N_{DoF}) \equiv \int_0^{z_i} f(z';N_{DoF})\,dz'$ distribution will be ***perfectly flat*** – *i.e.* it ***is*** the Uniform distribution $U(0{:}1)$ – the ***cumulative*** χ^2 P.D.F. $F(z_i;N_{DoF}) \equiv \int_0^{z_i} f(z';N_{DoF})\,dz'$ ***is*** ***uniformly distributed***!

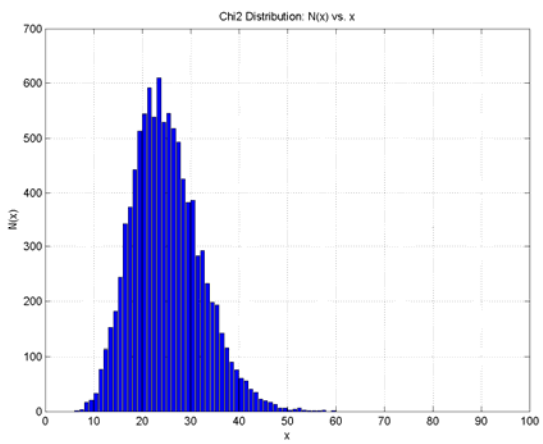
Put differently: If the χ^2 statistic is a ***random variable*** – distributed as the χ^2 P.D.F. for N_{DoF} ***degrees of freedom***, then the ***cumulative*** χ^2 P.D.F. $F(z_i;N_{DoF}) \equiv \int_0^{z_i} f(z';N_{DoF})\,dz'$ is ***also*** a ***random variable*** – distributed as its ***own*** P.D.F. – which ***is*** the ***Uniform distribution***, U(0,1)!

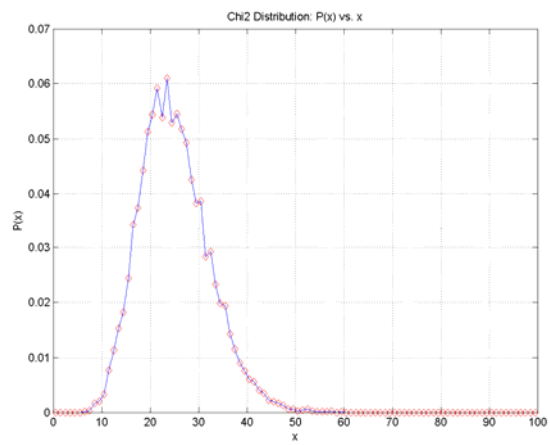
Then, since $p-value = CL_{upper}^{SS} \equiv 100\times\left[1 - F(z;N_{DoF})\right] = 100\times\left[1 - \int_0^{z} f(z';N_{DoF})\,dz'\right]$, the $p-value/CL_{upper}^{SS}$ distribution associated with the ensemble of a gazillion repetitions of the experiment will ***also*** be ***flat/uniformly distributed***, as $U(0{:}100\%)$!

The fact that the $p-value/CL_{upper}^{SS}$ distribution *should* be perfectly flat/uniform is a very useful / helpful tool, *e.g.* for testing/debugging Monte Carlo simulation programs…

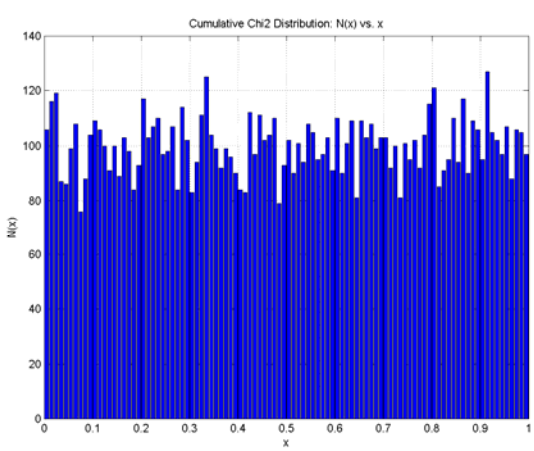
Another use of the (expected) flat/uniform $p-value/CL_{upper}^{SS}$ distribution is to make a ***cut*** on it somewhere between 0 and 100%, *e.g.* require $p-value/CL_{upper}^{SS} > 10.0\%$ in order to ***accept*** or ***reject*** a ***hypothesis*** that the data agrees *e.g.* with a theory prediction… the ***cut*** on the $p-value/CL_{upper}^{SS}$ is thus another example of a (formal) ***test statistic***.

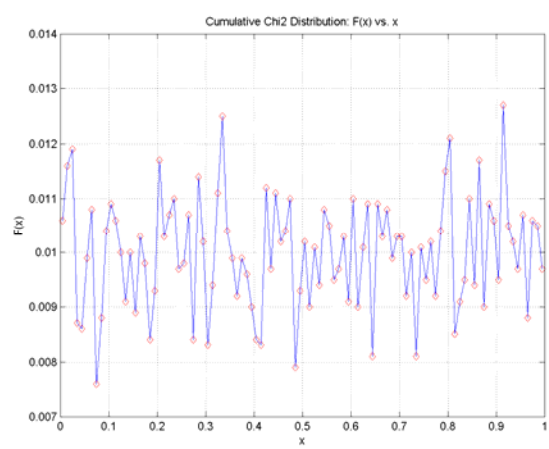
In the figures below, we show some Monte Carlo results associated with the repetition of an experiment 10,000× histogramming the resulting χ^2 distribution *e.g.* for P = 25 degrees of freedom. The ***true*** ***mean*** of this distribution is $\hat{\chi}^2 = P = 25$, with a ***variance*** of $\sigma^2_{\hat{\chi}^2} = 2P = 50$ and ***standard deviation*** of $\sigma_{\hat{\chi}^2} = \sqrt{\sigma^2_{\hat{\chi}^2}} = \sqrt{2P} = \sqrt{50} \simeq 7.07$. The ***maximum*** (*i.e.* the ***mode*** = ***most probable value***) occurs at $\chi^2_{\max} = P - 2 = 25 - 2 = 23$. The LHS plot below is a histogram of the χ^2 distribution for 10K repetitions of the experiment. We then convert this plot into a χ^2 probability distribution function (P.D.F.) $f\left(\chi^2; N_{DoF}\right)$ by dividing the contents of each histogram bin by the total # of events, which is shown in the RHS plot below:





A histogram of the resulting ***cumulative*** χ^2 P.D.F. $F\left(\chi^2 \le \chi_i^2; N_{DoF}\right) \equiv \int_0^{\chi_i^2} f\left(\chi^2; N_{DoF}\right) d\chi^2$ for 10K repetitions of this experiment is shown in the LHS plot below. We can again turn this into a ***cumulative*** χ^2 P.D.F. by dividing the contents of each histogram bin by the total # of events, as shown in the RHS plot below. Note that the ***cumulative*** χ^2 P.D.F. $F\left(\chi^2 \le \chi_i^2; N_{DoF}\right) \equiv \int_0^{\chi_i^2} f\left(\chi^2; N_{DoF}\right) d\chi^2$ ***is*** indeed distributed as the ***Uniform*** P.D.F. $U(0,1)$.





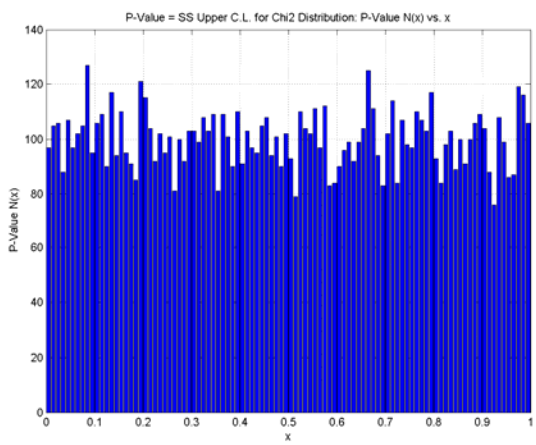
Since the ***cumulative*** χ^2 P.D.F. $F\left(\chi^2 \le \chi_i^2; N_{DoF}\right) \equiv \int_0^{\chi_i^2} f\left(\chi^2; N_{DoF}\right) d\chi^2$ ***is*** indeed distributed as the ***Uniform*** P.D.F. $U(0,1)$, then $1-F\left(\chi^2 \le \chi_i^2; N_{DoF}\right) = 1-\int_0^{\chi_i^2} f\left(\chi^2; N_{DoF}\right) d\chi^2$ ***is*** also distributed as the ***Uniform*** P.D.F. $U(0,1)$, and hence:

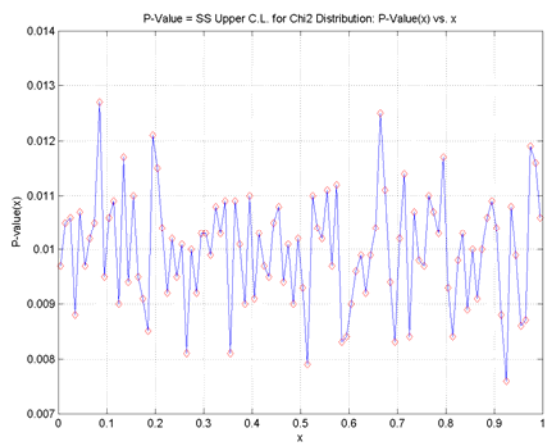
$$Prob\left(\chi^2 > \chi_i^2; N_{DoF}\right) = p-value = CL_{upper}^{SS} = \int_{\chi_i^2}^{\infty} f\left(\chi^2; N_{DoF}\right) d\chi^2$$
$$= 1-F\left(\chi^2 \le \chi_i^2; N_{DoF}\right) = 1-\int_0^{\chi_i^2} f\left(\chi^2; N_{DoF}\right) d\chi^2$$

is also distributed as the ***Uniform*** P.D.F. $U(0,1)$ as shown in the figures below. The LHS (RHS) plot, respectively is the histogram (corresponding P.D.F.) of

$$Prob\left(\chi^2 > \chi_i^2; N_{DoF}\right) = p-value = CL_{upper}^{SS} = \int_{\chi_i^2}^{\infty} f\left(\chi^2; N_{DoF}\right) d\chi^2$$
$$= 1-F\left(\chi^2 \le \chi_i^2; N_{DoF}\right) = 1-\int_0^{\chi_i^2} f\left(\chi^2; N_{DoF}\right) d\chi^2$$

for 10K repetitions of this experiment:





The uniformly-distributed behavior $U(0,1)$ of the ***cumulative*** P.D.F. $F\left(\chi^2 \le \chi_i^2; N_{DoF}\right)$ for the χ^2 distribution for a gazillion repetitions of an experiment is ***not*** unique/specific ***only*** to the χ^2 distribution – it is in fact a ***universal*** ***feature*** of ***all*** ***continuous*** P.D.F.'s and their associated ***cumulative*** P.D.F.'s – e.g. Gaussian/normal, exponential, log-normal, the uniform distribution itself, *etc*. For the ***discrete*** P.D.F.'s such as the binomial and the Poisson distributions, the ***cumulative*** P.D.F.'s are ***also*** ***discrete*** and thus, at first glance appear ***not*** to have uniform-distributed ***cumulative*** P.D.F.'s associated with them for gazillion repetitions of those binomial or Poisson-distributed experiments. However, if one ***smears out/averages*** over the histogram bin contents for the ***discrete*** distributions, then the resulting ***bin-averaged*** ***cumulative*** P.D.F.s ***are*** indeed ***also*** ***flat***/***uniform***. Plots of Matlab-based MC studies for various of these random distributions are posted on the P598AEM Software web page at:

http://courses.physics.illinois.edu/phys598aem/598aem_sw.html

<u>Pearson's</u> χ^2 <u>Test for Histograms</u>:

Suppose one has a ***histogram*** consisting of P bins of ***observed/measured*** values associated with a ***random variable*** *x*. The number of entries in the ith bin is $n_i(x_i)$, whereas *e.g.* a ***theory prediction*** exists *e.g.* with N $\underline{\lambda}$-parameters, such that the ***true mean/average*** number of entries ***expected/predicted*** is $\nu_i(x_i;\underline{\lambda})$. The most commonly used ***goodness-of-fit test*** for ***histograms*** is based on Pearson's χ^2 statistic:

$$\boxed{\chi^2_{Pearson}(\underline{\lambda}) \equiv \sum_{i=1}^{P} \frac{\left(n_i(x_i) - \nu_i(x_i;\underline{\lambda})\right)^2}{\nu_i(x_i;\underline{\lambda})}}$$

If the histogram bin data $\underline{n}(\underline{x}) = \left(n_1(x_1), n_2(x_2), \ldots, n_P(x_P)\right)$ are ***<u>Poisson-distributed</u>*** with ***mean*** values $\underline{\nu}(\underline{x}) = \left(\nu_1(x_1), \nu_2(x_2), \ldots, \nu_P(x_P)\right)$ **.and.** if the number of entries $n_i(x_i)$ in ***each*** histogram bin are not ***<u>too</u>*** small (*e.g.* $n_i(x_i) \gtrsim 5$), then it can be shown that Pearson's χ^2 statistic will indeed be distributed as a true χ^2 for $P-N$ degrees of freedom. This result holds regardless of the physical type/kind of $n_i(x_i)$ variable. Pearson's χ^2 test is thus said to be "***<u>distribution-free</u>***". Note that the $n_i(x_i) \gtrsim 5$ restriction on the number of entries in each histogram bin is equivalent to the requirement that the $n_i(x_i)$ be ***approximately*** Gaussian-distributed. Note further that a Poisson-distributed variable with ***mean*** ν_i also has ***variance*** ν_i and thus has a ***standard deviation*** of $\sigma_{\nu_i} = \sqrt{\nu_i}$. Thus, Pearson's χ^2 statistic is the sum of the squares of deviations between observed and expected values of the histogram bin content, measured in units of the ***<u>expected</u>*** (not ***<u>measured</u>***) ***standard deviations***, $\sigma_{\nu_i} = \sqrt{\nu_i}$.

Pearson's χ^2 statistic is also based on the implicit assumption that the total number of entries filling the histogram is ***<u>not</u> fixed***, *i.e.* such that $n_{tot} = \sum_{i=1}^{P} n_i(x_i)$ is ***<u>itself</u>*** a ***Poisson-distributed variable***.

If we ***<u>instead</u>*** regard $n_{tot} = \sum_{i=1}^{P} n_i(x_i)$ as being ***<u>fixed</u>/<u>constant</u>*** (*i.e.* only a ***<u>single</u>*** experiment has been carried out, not a gazillion repetitions of the experiment), then the $n_i(x_i)$ histogram bin data can be regarded as being ***<u>multinomially-distributed</u>***, with probabilities $p_i(x_i;\underline{\lambda}) = \nu_i(x_i;\underline{\lambda})/n_{tot}$. The relevant χ^2 statistic in ***<u>this</u>*** situation is then:

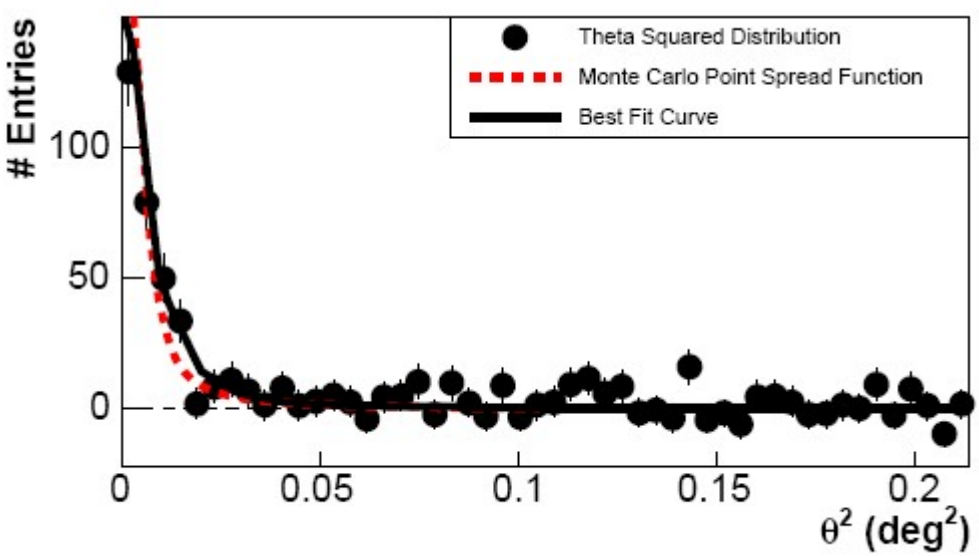
$$\boxed{\chi^2_{\substack{multi-\\nomial}}(\underline{\lambda}) \equiv \sum_{i=1}^{P} \frac{\left(n_i(x_i) - n_{tot}\, p_i(x_i;\underline{\lambda})\right)^2}{n_{tot}\, p_i(x_i;\underline{\lambda})}}$$

In the limit of a ***large*** number of entries each of the histogram bins, this multinomial χ^2 statistic is distributed as a true χ^2 for $P-N-1$ ***degrees of freedom***.

Below is an example of the potential use of Pearson's Poisson-based and/or the Multinomial-based χ^2 statistic – to compare histogrammed VHE (TeV) angle-squared gamma-ray data with angular resolution expectations associated with aVHE gamma ray source of angular size ~ 2′ (rms). The gamma rays emanating from the source, called HESS J1813-178 were subsequently found to be a supernova remnant located ~ in the plane of our own Milky Way galaxy.

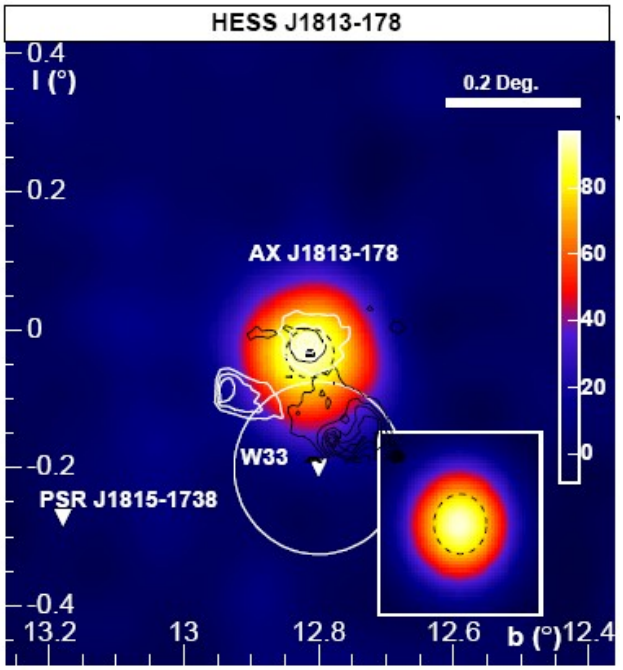
In a survey of the central region of the Milky Way with the H.E.S.S. Cherenkov telescopes eight previously unknown sources of very high energy (VHE) gamma rays were discovered. For two of these sources, HESS J1813-178 and HESS J1614-518, no counterparts in other wavelength regimes were found in the literature. In many sources of high-energy radiation, gamma ray production is related to the acceleration of electrons to high energies. These electrons will emit synchrotron X-rays and radio waves. A source not visible in X-rays or radio could either be a proton accelerator, or be immersed in a rather small magnetic field, reducing losses by synchrotron radiation.

Previously, two sources of VHE gamma rays without a counterpart were known, the HEGRA-discovered J2032+4130 and HESS J1303-631 . HESS J1813-178 is located within a fraction of a degree from the Galactic plane and is slightly extended, with about 2' (rms) angular size, as shown in the figure below. Therefore, it is most probably a Galactic object. The flux of VHE gamma rays is about 6% of the flux from the Crab nebula, and the energy spectrum extends to multi-TeV energies.

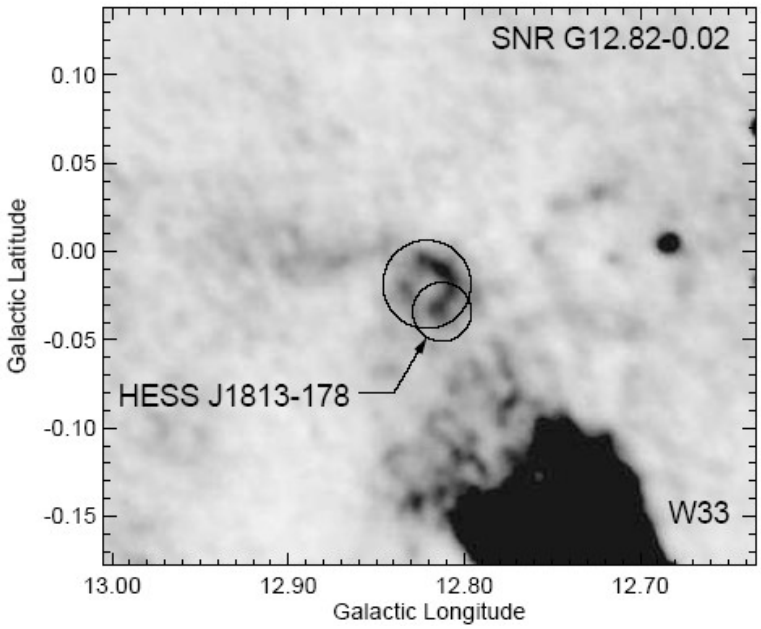


The angular distribution of VHE (TeV) gamma rays relative to the fitted source position. The red dashed line indicates the experimental resolution, the full black line the best fit obtained for an intrinsic source size of 2' rms.

The figure below shows the new TeV gamma-ray source HESS J1813-178 and its surrounding field of view. The inset illustrates the H.E.S.S. "beam size" governed both by the experimental resolution and a certain amount of smoothing applied to the image. The white contours show the smoothed X-ray count map obtained by the ASCA X-ray satellite and the black contours show 20 cm radio emission as observed by the VLA radio telescope.

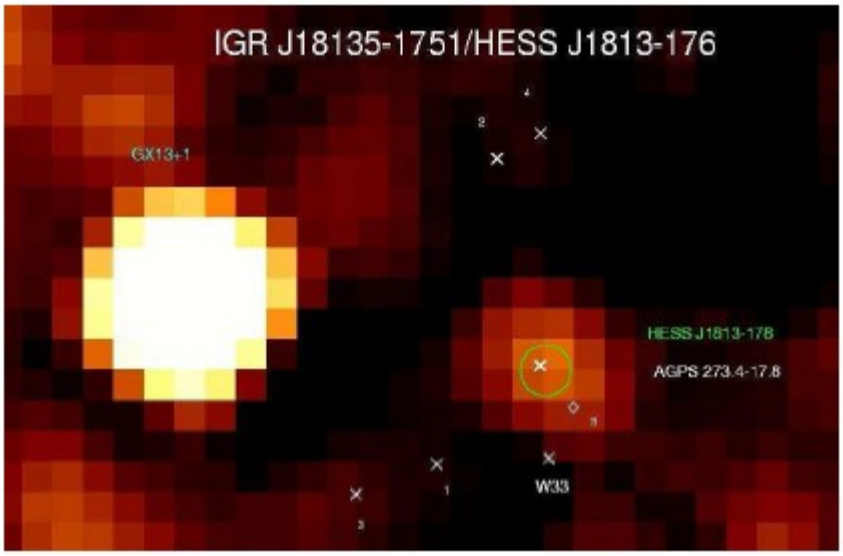


Shortly after the discovery of HESS J1813-178 was published, counterparts were located in existing but unpublished multi-wavelength data. At the location of HESS J1813-178, X-ray emission is seen in ASCA data. The ASCA X-ray source is termed AX J1813-178 or also AGPS273.4-17.8, the latter referring to the ASCA X-ray scan of the Galactic plane. In radio data, two experimental groups located a shell-type supernova remnant where a section of the shell coincides with the gamma-ray source, as shown in the figure below:



VLA 20 cm radio image of the region of the TeV source HESS J1813-178, showing the TeV source location and the newly discovered supernova remnant G18.82-0.02 (larger circle). The bright HII region W33 is seen at the lower right.

Data from the INTEGRAL satellite show in the 20-100 KeV range a soft gamma-ray source at the same location, as shown in the figure below.



INTEGRAL 20-100 keV soft gamma-ray image of the source region, with the strong source IGE J18135-1751 and a fainter source coincident with the TeV source, which is indicated as a green circle.

The distance to the supernova shell is estimated to be more than 4 kpc, from the size of the shell, estimated an age between 285 and 2500 years. Based on the available data, it is hard to identity the exact source mechanism - particle acceleration in the supernova shock wave as in RXJ 1713.7-3946, or emission from a pulsar wind nebula such as MSH 15-52 or the Crab Nebula. It is also possible that HESS J1813-178 may be part of a binary system, in analogy to the PSR B1259-63/SS2883 system. In either case, HESS J1813-178 is clearly more or less a "classical" TeV source, rather than a "dark accelerator", as initially thought.

References:

1.) “A new population of very high energy gamma-ray sources in the Milky Way”, the H.E.S.S. collaboration, F. Aharonian *et al*., Science 307 (2005) 1938-1942 and references therein.

2.)_“The H.E.S.S. survey of the Inner Galaxy in very high-energy gamma-rays”, F.A. Aharonian *et al*., Astrophysical Journal, 636 (2006) 777-797 and references therein.