$\mathcal{A}$ taste of statistics and applications to $\mathcal{X}$-ray spectral fitting
$\checkmark$ Normal error (Gaussian) distribution
$\rightarrow$ most important in statistical analysis of data, describes the distribution of random observations for many experiments
$\checkmark$ Poisson distribution
$\rightarrow$ generally appropriate for counting experiments related to random processes (e.g., radioactive decay of elementary particles)
$\checkmark$ Statistical tests: $\chi^{2}$ and F-test

Further details in the XSPEC presentation

## The Gaussian (normal error) distribution. I.

Averages of random variables (sufficiently large in number ) independently drawn from independent distributions converge in distribution to the normal
Casual errors are above and below the "true" (most "common") value
$\rightarrow$ bell-shape distribution if systematic errors are negligible


## The Gaussian probability function. II.

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

Probability Density Function (centered on $\mu$ )
$\mu=$ mean (expectation) value $\boldsymbol{\sigma}=$ standard deviation $\boldsymbol{\sigma}^{2}=$ variance


$$
e^{-x^{2} / 2 \sigma^{2}}
$$

$$
\text { function centered on } 0
$$

## The Gaussian probability function. III.



| 『 | 0 | 0,25 | 0,5 | 0,75 | 1,0 | 1,25 | 1,5 | 1,75 | 2,0 | 2,5 | 3,0 | 3,5 | 4,0 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $P(\%)$ | 0 | 20 | 38 | 55 | 68 | 79 | 87 | 92 | 95,4 | 98,8 | 99,7 | 99,95 | 99,99 |


$F(x)=\frac{1}{\sqrt{2 \pi} \sigma} \int_{-\infty}^{x} e^{-(x-\mu)^{2} / 2 \sigma^{2}} d x \quad$ Cumulative Distribution Function

## The Poisson distribution

Describes experimental results where events are counted and the uncertainty is not related to the measurement but reflects the intrinsically casual behavior of the process (e.g., radioactive decay of particles, X-ray photons, etc.)

$$
\mathrm{P}(x)=e^{-\mu} \mu^{x} / x!\quad(x=0,1,2, \ldots)
$$

Probability of obtaining $x$ events when $\mu$ events are expected $x=o b s e r v e d$ number of events in a time interval (frequency of events)
average
number
of events

$$
\bar{x}=\sum_{x=0}^{\infty} x P(x)=\sum_{x=0}^{\infty} x e^{-\mu} \mu^{x} / x!=\mu
$$

$\rightarrow \mu=$ average number of expected events if the experiment is repeated many times

$$
\begin{array}{ll}
\sigma^{2}=\left\langle(x-\mu)^{2}\right\rangle= & \begin{array}{l}
\text { expectation value of the } \\
\text { square of the deviations }
\end{array} \\
=\sum_{x=0}^{\infty}(x-\mu)^{2} \frac{\mu^{x}}{x!} e^{-\mu}=\mu
\end{array}
$$


the Poisson distribution with average counts $=\mu$ has standard deviation $V_{\mu}$

## Example: $\mathrm{N}_{\text {counts }} \pm \sqrt{ } \mathrm{N}$

High $\mu$ : the Poisson distribution is approximated by the Gaussian distribution
defined by only one parameter $\mu$


## F-test

If two statistics following the $\chi^{2}$ distribution have been determined, the ratio of the reduced chi-squares is distributed according to the F distribution

$$
P_{f}\left(f ; v_{1}, v_{2}\right)=\frac{\chi_{1}^{2} / v_{1}}{\chi_{2}^{2} / v_{2}}
$$

## $\propto \Delta \chi^{2} / k$

with $\mathrm{k}=$ number of additional terms (parameters)

Example: Use the F-test to evaluate the improvement to a spectral fit due to the assumption of a different model, with additional terms
Conditions: (a) the simpler model is nested within the more complex model;
(b) the extra parameters have Gaussian distribution (not truncated by the parameter space boundaries)
$\rightarrow$ see the F-test tables for the corresponding probabilities (specific command in XSPEC)

## An application of the F-test within XSPEC

| Model phabs<1>*powerlaw<2> Source No.: 1 | Active/On |  |  |  |  |  |
| :---: | :---: | :--- | :--- | :--- | :--- | :--- |
| Model Model Component | Parameter | Unit | Value |  |  |  |
| Mar | comp |  |  |  |  |  |
| 1 | 1 | phabs | nH | $10^{\wedge} 22$ | $1.59000 \mathrm{E}-02$ | frozen |
| 2 | 2 | powerlaw | PhoIndex |  | 2.72811 | $+/-$ |
| 3 | 2 | powerlaw | norm |  | $1.51490 \mathrm{E}-04$ | $+/-$ |

## Using energies from responses.

## Model1

Chi-Squared =
Reduced chi-squared =
97.23 using 105 PHA bins
0.9440 for 103 degrees of freedom
Null hypothesis probability $=6.417127 \mathrm{e}-01$
Model phabs<1>(laor<2> + powerlaw<3>) Source No.: 1 Active/On
Model Model Component Parameter Unit Value
par comp

| 1 | 1 | phabs | nH | $10^{\wedge} 22$ | $1.59000 \mathrm{E}-02$ | frozen |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 2 | 2 | laor | lineE | keV | 5.23582 | $+/-0.0$ |
| 3 | 2 | laor | Index |  | 3.00000 | frozen |
| 4 | 2 | laor | Rin(G) |  | 1.23500 | frozen |
| 5 | 2 | laor | Rout(G) |  | 400.000 | frozen |
| 6 | 2 | laor | Incl | deg | 30.0000 | frozen |
| 7 | 2 | laor | norm |  | $6.83065 \mathrm{E}-06$ | +- |
| 8 | 3 | powerlaw | PhoIndex |  | 2.77137 | $+/-0$ |
| 9 | 3 | powerlaw | norm |  | $1.48123 \mathrm{E}-04$ | $+/-0.0$ |

Using energies from response Model1+extra component

Chi-Squared =
Reduced chi-squared =
90.84 using 105 PHA bins.
0.8994 for $\quad 101$ degrees of freedom Null hypothesis probability $=7.557789 \mathrm{e}-01$
Current data and model not fit loat. F value $\Rightarrow$ low significance Weighting mothod. standerd 102 of the added component
F statistic value $=3.53567$ and probability 0.0327981



Fit (2) = Fit (1) + one component
xspec $>$ ftest $X^{2}$ (best fit) dof (best fit) $X^{2}$ (previous fit) dof (previous fit)
xspec> ftest $90.810197 .2103 \rightarrow$ ftest $=3.55 \rightarrow$ prob=0.0328

$$
\begin{aligned}
& \quad F_{t}=\left(\frac{\chi^{2}(d o f)-\chi^{2}(d o f-k)}{d o f-(d o f-k)}\right) /\left(\chi^{2}(d o f-k) /(d o f-k)\right)= \\
& =\left(\Delta \chi^{2} / k\right) / \chi_{\nu}^{2} \\
& \text { Ex: } \chi^{2}(103)=97.23 \\
& \chi^{2}(101)=90.84 \\
& \rightarrow \Delta \chi^{2}=6.39, k=2 \rightarrow F_{t}=(6.39 / 2) /(90.84 / 101)=3.55
\end{aligned}
$$

$F_{t}$ follows the $F$ distribution with $v_{1}=k=\Delta$ (dof) and $v_{2}=\operatorname{dof}-k(-1)$
Search in the F-distribution tables for the probability of the null hypothesis $\left(\mathrm{H}_{0}\right)$ for $v_{1}=2$ and $v_{2}=100$



