and it maps the N-vector $\boldsymbol{x}$ into our parameters $\boldsymbol{\alpha}$

$$\boldsymbol{\alpha} = G^+ \boldsymbol{x}. \tag{13.290}$$

The product $G^+ G = I_M$ is the $M \times M$ identity matrix, while

$$G\, G^+ = P \tag{13.291}$$

is an $N \times N$ projection operator (exercise 13.38) onto the $M \times M$ subspace for which $G^+ G = I_M$ is the identity operator. Like all projection operators, P satisfies $P^2 = P$.

13.20 Karl Pearson's Chi-Squared Statistic

The argument of the exponential (13.284) in $P(\boldsymbol{x})$ is (the negative of) Karl Pearson's chi-squared statistic (Pearson, 1900)

$$\chi^2 \equiv \sum_{j=1}^{N} \frac{(x_j - f(t_j;\boldsymbol{\alpha}))^2}{2\sigma^2}. \tag{13.292}$$

When the function $f(t;\boldsymbol{\alpha})$ is linear (13.286) in $\boldsymbol{\alpha}$, the N-vector $f(t_j;\boldsymbol{\alpha})$ is $f = G\,\boldsymbol{\alpha}$. Pearson's χ^2 then is

$$\chi^2 = (\boldsymbol{x} - G\,\boldsymbol{\alpha})^2/2\sigma^2. \tag{13.293}$$

Now (13.290) tells us that $\boldsymbol{\alpha} = G^+ \boldsymbol{x}$, and so in terms of the projection operator $P = G\,G^+$, the vector $\boldsymbol{x} - G\,\boldsymbol{\alpha}$ is

$$\boldsymbol{x} - G\,\boldsymbol{\alpha} = \boldsymbol{x} - G\,G^+ \boldsymbol{x} = \left(I - G\,G^+\right) \boldsymbol{x} = (I - P)\,\boldsymbol{x}. \tag{13.294}$$

So χ^2 is proportional to the squared length

$$\chi^2 = \tilde{\boldsymbol{x}}^2/2\sigma^2 \tag{13.295}$$

of the vector

$$\tilde{\boldsymbol{x}} \equiv (I - P)\,\boldsymbol{x}. \tag{13.296}$$

Thus if the matrix G has rank M, and the vector $\boldsymbol{x}$ has N independent components, then the vector $\tilde{\boldsymbol{x}}$ has only $N - M$ independent components.

Example 13.17 (Two Position Measurements) Suppose we measure a position twice with error σ, get x_1 and x_2, and choose $G^{\mathsf{T}} = (1, 1)$. Then the single parameter α is their average $\alpha = (x_1 + x_2)/2$, and χ^2 is

$$\begin{aligned}\chi^2 &= \left\{[x_1 - (x_1+x_2)/2]^2 + [x_2 - (x_1+x_2)/2]^2\right\} \Big/ \, 2\sigma^2 \\ &= \left\{[(x_1 - x_2)/2]^2 + [(x_2 - x_1)/2]^2\right\} \Big/ \, 2\sigma^2 \\ &= \left[(x_1 - x_2)/\sqrt{2}\right]^2 /2\sigma^2. \end{aligned} \tag{13.297}$$

Thus instead of having two independent components x_1 and x_2, χ^2 just has one $(x_1 - x_2)/\sqrt{2}$. □

We can see how this happens more generally if we use as basis vectors the $N - M$ orthonormal vectors $|j\rangle$ in the kernel of P (that is, the $|j\rangle$'s annihilated by P)

$$P|j\rangle = 0 \quad 1 \le j \le N - M \tag{13.298}$$

and the M that lie in the range of the projection operator P

$$P|k\rangle = |k\rangle \quad N - M + 1 \le k \le N. \tag{13.299}$$

In terms of these basis vectors, the N-vector $\boldsymbol{x}$ is

$$\boldsymbol{x} = \sum_{j=1}^{N-M} x_j|j\rangle + \sum_{k=N-M+1}^{N} x_k|k\rangle \tag{13.300}$$

and the last M components of the vector $\tilde{\boldsymbol{x}}$ vanish

$$\tilde{\boldsymbol{x}} = (I - P)\,\boldsymbol{x} = \sum_{j=1}^{N-M} x_j|j\rangle. \tag{13.301}$$

Example 13.18 (N position measurements) Suppose the N values of x_j are the measured values of the position $f(t_j;\alpha) = x_j$ of some object. Then $M = 1$, and we choose $G_{j1} = g_1(t_j) = 1$ for $j = 1, \ldots, N$. Now $G^{\mathsf{T}}G = N$ is a 1×1 matrix, the number N, and the parameter α is the mean $\overline{x}$

$$\alpha = G^{+}\,\boldsymbol{x} = \left(G^{\mathsf{T}}\,G\right)^{-1} G^{\mathsf{T}}\,\boldsymbol{x} = \frac{1}{N}\sum_{j=1}^{N} x_j = \overline{x} \tag{13.302}$$

of the N position measurements x_j. So the vector $\tilde{\boldsymbol{x}}$ has components $\tilde{x}_j = x_j - \overline{x}$ and is orthogonal to $G^{\mathsf{T}} = (1, 1, \ldots, 1)$

$$G^{\mathsf{T}}\tilde{\boldsymbol{x}} = \left(\sum_{j=1}^{N} x_j\right) - N\overline{x} = 0. \tag{13.303}$$

The matrix G^{T} has rank 1, and the vector $\tilde{\boldsymbol{x}}$ has $N-1$ independent components. □

Suppose now that we have determined our M parameters $\boldsymbol{\alpha}$ and have a theoretical fit

$$x = f(t;\boldsymbol{\alpha}) = \sum_{k=1}^{M} g_k(t)\,\alpha_k \tag{13.304}$$

which when we apply it to N measurements x_j gives χ^2 as

$$\chi^2 = (\tilde{\boldsymbol{x}})^2/2\sigma^2. \tag{13.305}$$

How good is our fit?

A χ^2 distribution with $N-M$ **degrees of freedom** has by (13.202) mean

$$E[\chi^2] = N - M \tag{13.306}$$

and variance

$$V[\chi^2] = 2(N-M). \tag{13.307}$$

So our χ^2 should be about

$$\chi^2 \approx N - M \pm \sqrt{2(N-M)}. \tag{13.308}$$

If it lies within this range, then (13.304) is a good fit to the data. But if it exceeds $N-M+\sqrt{2(N-M)}$, then the fit isn't so good. On the other hand, if χ^2 is less than $N-M-\sqrt{2(N-M)}$, then we may have used too many parameters or overestimated σ. Indeed, by using N parameters with $G\,G^{+} = I_N$, we could get $\chi^2 = 0$ every time.

The probability that χ^2 exceeds χ_0^2 is the integral (13.201)

$$\mathrm{Pr}_n(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} P_n(\chi^2/2)\,d\chi^2 = \int_{\chi_0^2}^{\infty} \frac{1}{2\Gamma(n/2)}\left(\frac{\chi^2}{2}\right)^{n/2-1} e^{-\chi^2/2} d\chi^2 \tag{13.309}$$

in which $n = N - M$ is the number of data points minus the number of parameters, and $\Gamma(n/2)$ is the gamma function (5.102, 4.62). So an M-parameter fit to N data points has only a chance of ϵ of being good if its χ^2 is greater than a χ_0^2 for which $\mathrm{Pr}_{N-M}(\chi^2 > \chi_0^2) = \epsilon$. These probabilities $\mathrm{Pr}_{N-M}(\chi^2 > \chi_0^2)$ are plotted in Fig. 13.6 for $N-M = 2$, 4, 6, 8, and 10. In particular, the probability of a value of χ^2 greater than $\chi_0^2 = 20$ respectively is 0.000045, 0.000499, 0.00277, 0.010336, and 0.029253 for $N-M = 2$, 4, 6, 8, and 10.

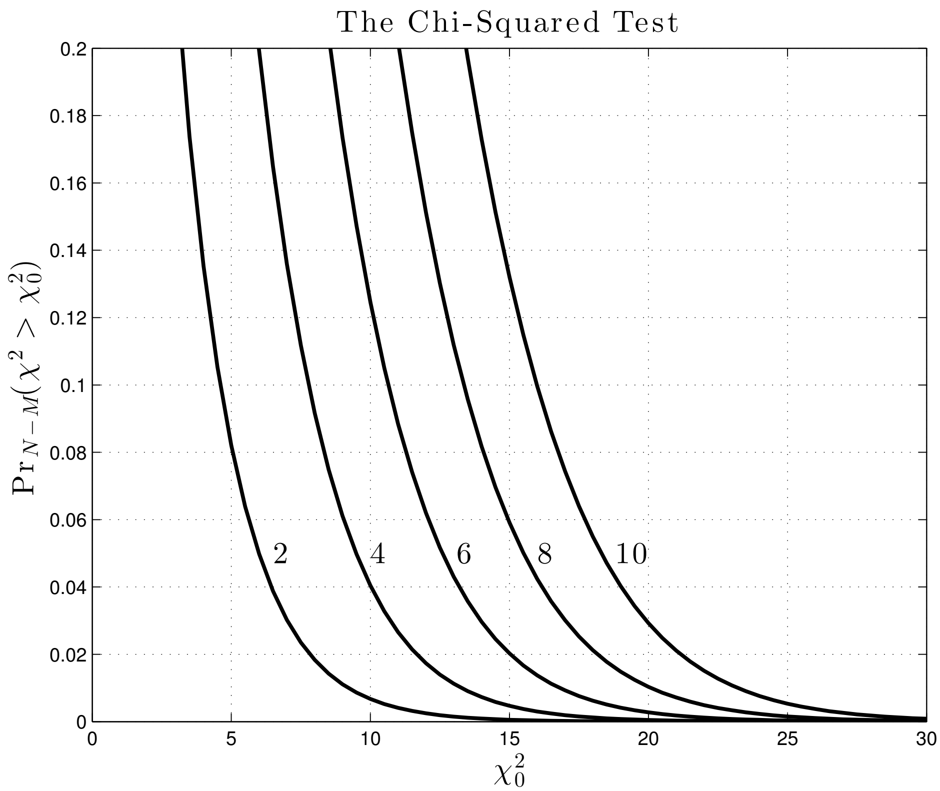

Figure 13.6 The probabilities $\Pr_{N-M}(\chi^2 > \chi_0^2)$ are plotted from left to right for $N - M = 2$, 4, 6, 8, and 10 degrees of freedom as functions of χ_0^2.

13.21 Kolmogorov's Test

Suppose we want to use a sequence of N measurements x_j to determine the probability distribution that they come from. Our empirical probability distribution is

$$P_e^{(N)}(x) = \frac{1}{N} \sum_{j=1}^{N} \delta(x - x_j). \tag{13.310}$$

Our cumulative probability for events less than x then is

$$\Pr_e^{(N)}(-\infty, x) = \int_{-\infty}^{x} P_e^{(N)}(x')\, dx' = \int_{-\infty}^{x} \frac{1}{N} \sum_{j=1}^{N} \delta(x' - x_j)\, dx'. \tag{13.311}$$