5.6 Several datasets, one test

The significance level \( p \) is an integral of a probability distribution function \( f \):

\[
\int_{\alpha}^{\infty} f(x) \, dx = p
\]

where \( \alpha \) is a critical value. For a particular set of data, \( \alpha \) is a statistic and \( p \) is random. Applying the change of variable rule

\[
\text{prob}(\alpha) \, d\alpha = \text{prob}(p) \, dp
\]

and remembering that \( f \) is the probability distribution of \( \alpha \), it follows that \( p \) is uniformly distributed between zero and one. (Notice this will not be true if the null hypothesis isn’t true, as then \( f \) will not be the probability distribution of \( \alpha \).)

Now

\[
\log W = \sum_{i=1}^{n} \log p_i
\]

is a sum of \( n \) random variables; \( \log p_i = u \) is distributed like \( e^{-u} \) (for \( u < 0 \)).

To find the distribution of the sum, we need the Fourier transform of this; it is proportional to \( 1/(k-i) \). The convolution theorem tells us that a sum of \( n \) terms will have a transform like \( 1/(k-i)^n \). This transform can be inverted for integer \( n \) (see tables, or the indispensable MATHEMATICA) and yields the required distribution of \( x = -\log W \). This is

\[
\frac{1}{\Gamma(n)} x^{n-1} e^{-x}.
\]

The results on the mean and variance follow easily from direct integrations. To see the Gaussian form, expand the log of the distribution function in a Taylor series around \( n \), to second order.