## THE ROYAL STATISTICAL SOCIETY

## 2007 EXAMINATIONS - SOLUTIONS

## GRADUATE DIPLOMA

## STATISTICAL THEORY AND METHODS

## PAPER II

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Note. In accordance with the convention used in the Society's examination papers, the notation log denotes logarithm to base e. Logarithms to any other base are explicitly identified, e.g. $\log _{10}$.
(a) The distribution of $\bar{X}$ is $\mathrm{N}\left(\mu, \sigma^{2} / n\right)$. The distribution of $\bar{Y}$ is $\mathrm{N}\left(\rho \mu, \rho^{2} \sigma^{2} / n\right)$, and therefore the distribution of $\bar{Y} / \rho$ is $\mathrm{N}\left(\mu, \sigma^{2} / n\right) . \bar{X}$ and $\bar{Y}$ are independent.

A random variable is a pivotal quantity for particular parameters if its distribution is independent of the parameters.

We have that $\bar{X}-(\bar{Y} / \rho) \sim \mathrm{N}\left(0,2 \sigma^{2} / n\right)$, so $\frac{\bar{X}-(\bar{Y} / \rho)}{\sqrt{2 \sigma^{2} / n}} \sim \mathrm{~N}(0,1)$; therefore this is a pivotal quantity.
(b) (i) Let $U$ have pdf $f(u)=\theta^{-1} u^{(1-\theta) / \theta} \quad$ (for $0<u<1$ ), and let $T=-\theta^{-1} \log U$.

By standard transformation results, the pdf of $T$ is given by $g(t)=f(u)\left|\frac{d u}{d t}\right|$.
Here $t=-\theta^{-1} \log u$, so $u=e^{-\theta t}$ and $\left|\frac{d u}{d t}\right|=\left|-\theta e^{-\theta t}\right|$.
$\therefore g(t)=\theta^{-1}\left(e^{-\theta t}\right)^{(1-\theta) / \theta}\left|-\theta e^{-\theta t}\right|=\frac{1}{\theta} e^{-t} e^{\theta t} \theta e^{-\theta t}=e^{-t}$,
and this is valid for $t>0$ (since $0<u<1 \Rightarrow-\log u>0)$.

$$
\begin{aligned}
& E(T)=\int_{0}^{\infty} t e^{-t} d t=\left[-t e^{-t}\right]_{0}^{\infty}+\int_{0}^{\infty} e^{-t} d t=0+\left[-e^{-t}\right]_{0}^{\infty}=1 . \\
& E\left(T^{2}\right)=\int_{0}^{\infty} t^{2} e^{-t} d t=\left[-t^{2} e^{-t}\right]_{0}^{\infty}+2 \int_{0}^{\infty} t e^{-t} d t=0+2 E(T)=2 . \\
& \therefore \operatorname{Var}(T)=2-1^{2}=1 .
\end{aligned}
$$

(ii) $V$ is the sum of $n$ independent exponential random variables each with mean 1 , so $V$ has the gamma distribution with parameters $n$ and 1 . Since $\theta$ is not involved, $V$ is a pivotal quantity.

## Solution continued on next page

(iii) For $n$ large, $V$ is approximately distributed as $\mathrm{N}(n, n)$, and so to this level of approximation $P\left(-1.96<\frac{V-n}{\sqrt{n}}<1.96\right)=0.95$,
i.e. $\quad P(n-1.96 \sqrt{n}<V<n+1.96 \sqrt{n})=0.95$,
i.e. $\quad P\left(n-1.96 \sqrt{n}<-\frac{1}{\theta} \sum_{i=1}^{n} \log U_{i}<n+1.96 \sqrt{n}\right)=0.95$.

Hence the required approximate $95 \%$ confidence interval for $\theta$ is given by

$$
\frac{-\sum \log u_{i}}{n+1.96 \sqrt{n}}<\theta<\frac{-\sum \log u_{i}}{n-1.96 \sqrt{n}}
$$

(i) Let $\boldsymbol{X}=\left(X_{1}, X_{2}, \ldots, X_{n}\right)$ represent the data; then, if $T$ is sufficient for the parameter $\theta$, the distribution of $\boldsymbol{X}$ given $T$ does not involve $\theta$.

The factorisation theorem states that $T$ is sufficient if and only if the likelihood $L(\boldsymbol{x}, \theta)$ can be written as a product $L(\boldsymbol{x}, \theta)=g(T, \theta) h(\boldsymbol{x})$.
(ii) From the factorisation theorem, maximising the likelihood $L(\boldsymbol{x}, \theta)$ with respect to $\theta$ is equivalent to maximising $g(T, \theta)$ with respect to $\theta$. Since $g$ depends on the data only through $T$, the maximum likelihood estimator depends on the data only through $T$. The maximum likelihood estimator must therefore be a function of the sufficient statistic $T$.
(iii) As an example, consider the one-parameter gamma distribution with pdf $f(x, \theta)=x^{\theta-1} e^{-x} / \Gamma(\theta)$.

Given a sample of $n$ independent observations, $\Pi X_{i}$ is sufficient for $\theta$ [in the examination, candidates could quote this as a standard result or derive it by factorising the likelihood]. However, the mean of the distribution is $\theta$, so the method of moments estimator of $\theta$ is $\bar{X}$.
(iv) The likelihood for a sample of $n$ independent observations is

$$
\begin{array}{r}
L(\boldsymbol{x}, \theta)=\frac{1}{(\theta \sqrt{2 \pi})^{n}} \frac{1}{\prod_{i=1}^{n} x_{i}} \exp \left\{-\frac{1}{2 \theta^{2}} \sum\left(\log x_{i}\right)^{2}\right\} \\
=\frac{1}{\theta^{n}} \exp \left\{-\frac{1}{2 \theta^{2}} \sum\left(\log x_{i}\right)^{2}\right\} \cdot\left(\frac{1}{\sqrt{2 \pi}}\right)^{n} \frac{1}{\prod \longleftrightarrow x_{i}} \\
\longleftrightarrow g(T, \theta) \longrightarrow \boldsymbol{x}) \longrightarrow
\end{array}
$$

which shows that $\Sigma\left(\log x_{i}\right)^{2}$ is sufficient for $\theta$.

$$
p(x)=\theta(1-\theta)^{x}, \quad x=0,1,2, \ldots, \quad 0<\theta<1 .
$$

(i) The likelihood is $L=\prod_{i=1}^{n} \theta(1-\theta)^{x_{i}}$.
$\therefore \log L=n \log \theta+\sum_{i} x_{i} \log (1-\theta) . \quad \therefore \frac{d}{d \theta}(\log L)=\frac{n}{\theta}-\frac{\sum x_{i}}{1-\theta}$.

Setting this equal to zero gives $\hat{\theta} \sum x_{i}=n(1-\hat{\theta})$, i.e. $\hat{\theta}=\frac{n}{n+\Sigma x_{i}}=\frac{1}{1+\bar{x}}$. It may be verified (e.g. by considering the second derivative - see part (ii) below) that this is indeed a maximum, and so the maximum likelihood estimator of $\theta$ is $\hat{\theta}=\frac{1}{1+\bar{X}}$.

To find the method of moments estimator $\tilde{\theta}$, we first obtain $E[X]=\sum_{x=0}^{\infty} x \theta(1-\theta)^{x}=\theta(1-\theta) \theta^{-2}$ using the result quoted in the question.

Thus $E[X]=\frac{1-\theta}{\theta}=\frac{1}{\theta}-1$, and therefore $\tilde{\theta}$ satisfies $\bar{X}=\frac{1}{\tilde{\theta}}-1$ so that $\tilde{\theta}=\frac{1}{1+\bar{X}}$.
Thus, in this case, the maximum likelihood estimator is the same as the method of moments estimator.
(ii) Differentiating $\frac{d \log L}{d \theta}$ (see above) gives $\frac{d^{2} \log L}{d \theta^{2}}=-\frac{n}{\theta^{2}}-\frac{\sum x_{i}}{(1-\theta)^{2}}$.
$\therefore E\left(-\frac{d^{2} \log L}{d \theta^{2}}\right)=\frac{n}{\theta^{2}}+\frac{n(1-\theta)}{\theta(1-\theta)^{2}}=\frac{n}{\theta^{2}(1-\theta)}$.
Therefore the Cramér-Rao lower bound is $\frac{\theta^{2}(1-\theta)}{n}$, as required.

Solution continued on next page
(iii) In large samples, $\hat{\theta}$ may be taken as unbiased for $\theta$ and $\operatorname{Var}(\hat{\theta}) \approx \frac{\theta^{2}(1-\theta)}{n}$ (the Cramér-Rao lower bound), which we may estimate by $\frac{\hat{\theta}^{2}(1-\hat{\theta})}{n}$.

So a large-sample approximate $95 \%$ confidence interval for $\theta$ is given by

$$
\left(\hat{\theta}-1.96 \frac{\hat{\theta} \sqrt{1-\hat{\theta}}}{\sqrt{n}}, \hat{\theta}+1.96 \frac{\hat{\theta} \sqrt{1-\hat{\theta}}}{\sqrt{n}}\right) .
$$

(iv) For the lower limit of the confidence interval to be less than 0 , we would need the value of $1.96 \hat{\theta} \sqrt{(1-\hat{\theta}) / n}$ to exceed $\hat{\theta}$ itself. This means that we would require $1.96 \sqrt{1-\hat{\theta}}>\sqrt{n}$, i.e. $\hat{\theta}<1-\frac{n}{1.96^{2}}$. But we know that $\hat{\theta}$ must be between 0 and 1 , so this can only happen if $n$ is 1,2 or 3 .

Thus the comment has no validity provided the sample size is at least 4 . The comment is irrelevant because the confidence interval is a large-sample approximation and should certainly not be used for samples of size as small as 3 .
(i) When $\theta=1, L(T, \theta)=T+\frac{1}{T}-2$. In the examination, any reasonable sketch of this was accepted. The main properties are that $L \rightarrow \infty$ as $T \rightarrow 0 ; L \rightarrow \infty$ as $T \rightarrow \infty ; L=0$ for $T=1$, and otherwise $L>0 ;$ also, $\frac{d L}{d T}=1-\frac{1}{T^{2}}$. To save space, the sketch is not shown in this solution.
(ii) For $T=r \theta, L=r+\frac{1}{r}-2$.

By inspecting the form of the function, $T=\frac{\theta}{r}$ must give the same result.
Hence being wrong by a factor of (say) 2 in the estimate carries the same loss as being wrong by a factor of $1 / 2$.
(iii) $\quad X_{1}+X_{2}$ is the sum of two independent exponential random variables each with mean $\theta$, and so $X_{1}+X_{2}$ has the gamma distribution with parameters 2, $\theta$. That is, letting $Y$ denote $X_{1}+X_{2}$, we have that the pdf of $Y$ is $y e^{-y / \theta} / \theta^{2}$ (for $y>0$ ).

$$
\therefore E\left(\frac{1}{Y}\right)=\int_{0}^{\infty} \frac{1}{y} \cdot \frac{y}{\theta^{2}} e^{-y / \theta} d y=\frac{1}{\theta^{2}} \int_{0}^{\infty} e^{-y / \theta} d y=\frac{1}{\theta^{2}}\left[-\theta e^{-y / \theta}\right]_{0}^{\infty}=\frac{1}{\theta} .
$$

(iv) $\quad T=c \bar{X}=\frac{1}{2} c Y$.
$\therefore E(T)=c E(\bar{X})=c \theta \quad$ and $\quad E\left(\frac{1}{T}\right)=\frac{2}{c} E\left(\frac{1}{Y}\right)=\frac{2}{c \theta}$.
$\therefore E[L(T, \theta)]=c+\frac{2}{c}-2$.
$\therefore \frac{d E(L(T, \theta))}{d c}=1-\frac{2}{c^{2}}$, which equals zero when $c= \pm \sqrt{ } 2$, and here we must take $c=+\sqrt{ } 2$ as $c$ is required to be positive. To confirm that this gives a minimum of the expected loss, consider $\frac{d^{2} E(L(T, \theta))}{d c^{2}}=\frac{4}{c^{3}}$ which is $>0$ for $c=+\sqrt{ } 2$, so this value of $c$ does indeed give a minimum.
(i) On $H_{0}, P(X=x)=\frac{e^{-1}}{x!}$. On $H_{1}, P(X=x)=\frac{e^{-2} 2^{x}}{x!}$. (For $x=0,1,2, \ldots$ )

So the likelihoods are $L_{0}=\frac{e^{-n}}{\prod_{i=1}^{n} x_{i}!}$ and $L_{1}=\frac{e^{-2 n} 2^{\Sigma x_{i}}}{\prod_{i=1}^{n} x_{i}!}$. So $\frac{L_{1}}{L_{0}}=e^{-n} 2^{\Sigma x_{i}}$, and the Neyman-Pearson method requires us to reject $H_{0}$ in favour of $H_{1}$ when $\frac{L_{1}}{L_{0}}$ is large. This ratio is an increasing function of $\sum x_{i}$, so rejection happens if $\sum_{i=1}^{n} x_{i} \geq k$, for a suitable value of $k$.
(ii) Let $T=\Sigma X_{i}$. We have that $T \sim \mathrm{~N}(n \theta, n \theta)$, approximately.

On $H_{0}, \theta=1$ and so $T \sim \mathrm{~N}(n, n)$, approximately. The upper $1 \%$ point of $\mathrm{N}(0,1)$ is 2.3263 . So the criterion for Type I error requires that we reject $H_{0}$ if $\frac{T-n}{\sqrt{n}} \geq 2.3263$, i.e. if $T \geq n+2.3263 \sqrt{n}$.

On $H_{1}, \theta=2$ and so $T \sim \mathrm{~N}(2 n, 2 n)$, approximately. So the criterion for Type II error now requires that $P(T \geq n+2.3263 \sqrt{n} \mid T \sim \mathrm{~N}(2 n, 2 n))=0.99$, giving

$$
1-\Phi\left(\frac{n+2 \cdot 3263 \sqrt{n}-2 n}{\sqrt{2 n}}\right)=0.99 \text {. }
$$

$\therefore \frac{n+2.3263 \sqrt{n}-2 n}{\sqrt{2 n}}=-2.3263$, which gives $\sqrt{n}=2.3263(1+\sqrt{2})$.
Thus $n=\{2.3263(1+\sqrt{2})\}^{2}=31.54$, and we take $n=32$.
(iii) To test $H_{0}$ against $H_{1}{ }^{*}: \theta=\theta^{*}>1$, the likelihood ratio will be $e^{-n\left(\theta^{*}-1\right)}\left(\theta^{*}\right)^{\Sigma x_{i}}$ and the Neyman-Pearson method rejects $H_{0}$ for $\sum_{i=1}^{n} x_{i} \geq k$ as above. The form of the test is the same for any $\theta^{*}>1$, so the test is uniformly most powerful. As above, for a test of size approximately 0.01 we reject $H_{0}$ if $\Sigma x_{i} \geq n+$ $2.3263 \sqrt{ } n$. For the sketch, note that the power is (approximately) 0.01 for $\theta=1$ and 0.99 for $\theta=2$. The curve will follow the usual S-shape. To save space, the sketch is not shown in this solution.
(i) Let $L_{0 m}$ and $L_{1 m}$ be the likelihoods under $H_{0}$ and $H_{1}$ after $m$ observations have been sampled. Let $\lambda_{m}=L_{0 m} / L_{1 m}$.

For constants $k_{0}$ and $k_{1}$ to be determined (see below), the SPRT procedure is: continue sampling if $k_{0}<\lambda_{m}<k_{1}$; reject $H_{0}$ if $\lambda_{m} \leq k_{0}$; reject $H_{1}$ if $\lambda_{m} \geq k_{1}$.

The approximate values of $k_{0}$ and $k_{1}$ are as follows:

$$
k_{0} \approx \frac{\alpha}{1-\beta}, \quad k_{1} \approx \frac{1-\alpha}{\beta}
$$

Now let $z_{i}=\log \left[\frac{f_{0}\left(x_{i}\right)}{f_{1}\left(x_{i}\right)}\right]$ where $f_{j}\left(x_{i}\right)$ is the value of the pdf at observation $x_{i}$ under hypothesis $H_{j}$.

Wald's formulae are

$$
\begin{aligned}
& E\left(N \mid H_{0}\right) \approx \frac{\alpha \log \left(\frac{\alpha}{1-\beta}\right)+(1-\alpha) \log \left(\frac{1-\alpha}{\beta}\right)}{E\left(Z_{i} \mid H_{0}\right)}, \\
& E\left(N \mid H_{1}\right) \approx \frac{(1-\beta) \log \left(\frac{\alpha}{1-\beta}\right)+\beta \log \left(\frac{1-\alpha}{\beta}\right)}{E\left(Z_{i} \mid H_{1}\right)}
\end{aligned}
$$

(ii) If $n$ is the fixed sample size, usually $E\left(N \mid H_{0}\right)<n$ and $E\left(N \mid H_{1}\right)<n$, though occasionally there is a sample where $N>n$.

For $\theta_{0}<\theta<\theta_{1}, E(N)$ will be larger in the middle of the range than it is at the boundaries and $P(N>n \mid \theta)$ tends to increase for these intermediate values.
(iii) (a) $\lambda_{m}=\frac{\prod_{i=1}^{m} \frac{1}{\sqrt{2 \pi}} e^{-x_{i}^{2} / 2}}{\prod_{i=1}^{m} \frac{1}{2 \sqrt{2 \pi}} e^{-x_{i}^{2} / 8}}=2^{m} e^{-(3 / 8) \Sigma x_{i}^{2}}$.

For tests with error probabilities $\alpha$ and $\beta$ both approximately 0.05 , we take $k_{0}=1 / 19$ and $k_{1}=19$ in the method outlined in part (i).

We have $z_{i}=\log 2-\frac{3}{8} x_{i}^{2}$ and so

$$
\begin{aligned}
& E\left(Z_{i} \mid H_{0}\right)=\log 2-\frac{3}{8} \text { and } E\left(Z_{i} \mid H_{1}\right)=\log 2-\left(\frac{3}{8} \times 4\right)=\log 2-\frac{3}{2} . \\
\therefore & E\left(N \mid H_{0}\right)=\frac{0.9 \log 19}{\log 2-\frac{3}{8}}=8.33
\end{aligned}
$$

and $\quad E\left(N \mid H_{1}\right)=\frac{-0.9 \log 19}{\log 2-\frac{3}{2}}=3.28$.
(b) $\quad \lambda_{1}=2 e^{-\frac{3}{8}(1.6)^{2}}=0.766$, so continue sampling.
$\lambda_{2}=2^{2} e^{\left.-\frac{3}{8}(1.6)^{2}+(-0.9)^{2}\right)}=4 e^{-3 \times 3 \times 37 / 8}=1.130$, so continue sampling.
$\lambda_{3}=2^{3} e^{\left.-\frac{3}{8}(1.6)^{2}+(-0.9)^{2}+(-2.5)^{2}\right)}=8 e^{-3 \times 9.62 / 8}=0.217$, so continue sampling.
[ $k_{0}=\frac{1}{19}=0.053$ and $k_{1}=19$ are the critical points.]
(i) The likelihood is

$$
L(\mathbf{x} \mid \mu)=\left(\frac{1}{\sqrt{2 \pi}}\right)^{n} \exp \left(-\frac{\Sigma\left(x_{i}-\mu\right)^{2}}{2}\right) \propto \exp \left(\frac{2 \mu \Sigma x_{i}-n \mu^{2}}{2}\right)
$$

The density of the prior is

$$
\pi(\mu)=\frac{1}{\sqrt{2 \pi}} \exp \left(-\frac{(\mu-a)^{2}}{2}\right) \propto \exp \left(\frac{2 \mu a-\mu^{2}}{2}\right)
$$

Hence the density of the posterior is

$$
\begin{aligned}
\pi(\mu \mid \mathbf{x}) & \propto \pi(\mu) L(\mathbf{x} \mid \mu) \\
& \propto \exp \left(\frac{2 \mu\left(\Sigma x_{i}+a\right)-(n+1) \mu^{2}}{2}\right) \propto \exp \left\{-\frac{1}{2} \frac{\left(\mu-\frac{\Sigma x_{i}+a}{n+1}\right)^{2}}{\frac{1}{n+1}}\right\}
\end{aligned}
$$

which can be seen to be of the form $\exp \left(\frac{(x-A)^{2}}{2 B^{2}}\right)$ so that it is the kernel of the pdf of the Normal distribution $\mathrm{N}\left(\frac{\sum x_{i}+a}{n+1}, \frac{1}{n+1}\right)$, as required.
(ii) (a) This prior is a two-component Normal mixture, equally weighted: $\frac{1}{2}\{\mathrm{~N}(2,1)+\mathrm{N}(-2,1)\}$. The two component means are 4 units apart, so the tails of the mixture will look just as in the basic Normal components. There is a small overlap around 0 in which the shape will not be quite Normal. The pdf is symmetrical because each component is equally weighted (each has weight $1 / 2$ ).

In the examination, any reasonable sketch was accepted. To save space, a sketch is not shown in this solution.

## Solution continued on next page

(b) The likelihood $L(\mathbf{x}, \mu)$ is as in part (i).

The prior can be split into two parts, and these can be dealt with separately, giving results as in part (i), by setting $a_{1}=+2$ and $a_{2}=-2$ to obtain

$$
\pi(\mu) \propto \sum_{j=1}^{2} \exp \left(\frac{2 \mu a_{j}-\mu^{2}}{2}\right), \text { for } j=1,2 .
$$

This leads to

$$
\pi(\mu \mid \mathbf{x}) \propto \sum_{j=1}^{2} \exp \left\{-\frac{1}{2} \frac{\left(\mu-\frac{\Sigma x_{i}+a_{j}}{n+1}\right)^{2}}{\frac{1}{n+1}}\right\}
$$

The posterior is therefore an equal probability mixture of $\mathrm{N}\left(\frac{\Sigma x_{i}-2}{n+1}, \frac{1}{n+1}\right)$ and $\mathrm{N}\left(\frac{\Sigma x_{i}+2}{n+1}, \frac{1}{n+1}\right)$.

In the particular case in the question, these are $\mathrm{N}(0.86,1 / 100)$ and $\mathrm{N}(0.90,1 / 100)$.

Unlike the mixture that forms the prior, the component means here are very close and there is a large overlap in the centre: about $95 \%$ of the respective parts lie in the ranges $(0.66,1.06)$ and $(0.70,1.10)$. When the two components are combined, the result will look very like $\mathrm{N}(0.88,1 / 100)$ - note the small spread, the standard deviation is only 1/10.

Again, any reasonable sketch was accepted in the examination. A sketch is not shown here.

## (This solution is continued on the next page)

## (a) Kolmogorov-Smirnov test

Given a random sample from a population with a hypothesised specified cumulative distribution function $F(x)$, the sample (empirical) cumulative distribution function $S(x)$ is found as follows and compared with the specified $F(x)$. With $n$ observations in the sample, define $S(x)$ as $(1 / n) \times$ the number of sample values that are $\leq x$, for each observed $x$. If the sample values are arranged in order as $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, then $S(x)$ increases by $1 / n$ at each new value $x_{(j)}$. So the graph of $S(x)$ against $x$ is a step function.
Basically the test examines whether $F(x)$ is fairly close to $S(x)$, i.e. whether the graph of $F(x)$ is close to the step function $S(x)$. The maximum difference between $S(x)=j / n$ and the corresponding $F\left(x_{(j)}\right)$ is found, and referred to a special table of critical values for the usual significance levels. If this maximum difference is not significant, the hypothesis that that cdf is $F(x)$ is not rejected. This test is distribution-free, i.e. it can be used for any (fully specified) $F(x)$. It can be used for samples of any size, including small ones.

## Chi-squared test

This test, basically for the same purpose, needs a reasonable amount of data. The data have to be grouped into suitable intervals, a somewhat arbitrary process which can affect inferences. Once grouped, the frequencies in each interval are found and compared with the frequencies to be expected on the proposed hypothesis, using the formula $\Sigma\left(O_{i}-E_{i}\right)^{2} / E_{i}$ where $O_{i}$ and $E_{i}$ are the observed and expected frequencies in the $i$ th interval. The number of degrees of freedom is found by the usual rule (number of intervals -1 ) and the usual $\chi^{2}$ table is used. If Cochran's proposal to choose intervals with roughly equal expected frequencies is followed, the power of this test is fairly good, but skew distributions with long tails may cause problems (as do some discrete distributions).
The test is approximately distribution-free in large samples. The approximations inherent in it are usually close to being satisfied provided there are no "small" expected frequencies. The criterion that no $E_{i}$ may be less than 5 is often used, though this is often thought to be too restrictive.

## Situation with unknown parameter(s)

If the basic form of the hypothesised underlying distribution is known but with unspecified parameter(s) [e.g. it is Normal but its mean and variance are not known], the parameter(s) must be estimated from the sample data. $F(x)$ can then be constructed using the estimated value(s); this will be satisfactory if the sample is sufficiently large for the estimates to be reliable. The test procedure can then be carried through as before, but there is the problem of "over-fitting" since the data are being compared with an $F(x)$ which is automatically "closer" to the data by virtue of using the estimated parameter(s). The chi-squared test provides a built-in adjustment by reducing the number of degrees of freedom
by one for each parameter that is estimated; the test remains a satisfactory procedure to a good level of approximation. There is no real way to adjust the Kolmogorov-Smirnov test in general. If it is simply used as described above, with its standard tables, the procedure will be conservative, possibly substantially so. Special tables might be available for particular hypothesised distributions, but each distribution needs its own table.
(b) Suppose the two samples are of sizes $n_{1}$ and $n_{2}$ and it may be assumed that the underlying populations are Normally distributed with known variances $\sigma_{1}{ }^{2}$ and $\sigma_{2}{ }^{2}$. The means $\mu_{1}$ and $\mu_{2}$ may sensibly be used as location parameters and the null hypothesis will typically be $\mu_{1}=\mu_{2}$. A Normal-based parametric approach is to calculate the value of the test statistic

$$
z=\frac{\bar{x}_{1}-\bar{x}_{2}}{\sqrt{\frac{\sigma_{1}{ }^{2}}{n_{1}}+\frac{\sigma_{2}{ }^{2}}{n_{2}}}} \quad \text { (where } \bar{x}_{1} \text { and } \bar{x}_{2} \text { are the sample means) }
$$

and refer it to the null distribution $\mathrm{N}(0,1)$ in the usual way. If the samples are large, the Central Limit Theorem shows that this method can still be used as a good approximation. This remains the case when the population variances are unknown, estimating them by the sample variances on the basis that, in large samples, these estimates should be satisfactory. If the samples are small and the population variances are unknown, this method is not reliable; but, for the case where the underlying distributions are Normal, a similar method based on the $t$ distribution can be used.

A suitable rank-based test is Wilcoxon's. Here the responses are ranked in order from the smallest to the largest in a single ranking. The sum of the ranks corresponding to one of the samples is then calculated. This is referred to a table of critical values, under the hypothesis of no difference in the locationparameters of the underlying distributions but on the assumption that these distributions are otherwise identical.

In general, Normal-based tests are more powerful than rank-based tests if the underlying distributions are indeed Normal (or if the samples are sufficiently large that the Central Limit Theorem may be reliably invoked). However, rank-based tests are likely to be more robust to departures from the assumption of Normality or to the presence of outliers, and may be preferred for this reason. This is especially true if the samples are small. Rank-based tests are also often easy to use, in terms of calculating the value of the test statistic and referring it to tables of its critical values (further, easy-to-use Normal approximations to null distributions are available and are good even for only moderately large samples).

