POPULATION VALUE OF SPEARMAN'S MODIFIED FOOTRULE CORRELATION COEFFICIENT

Let (X_1, Y_1) be a standarized bivariate normal random variable with correlation coefficient r, and X an independent N(0,1) random variable. For Φ equal to the cumulative distribution function of X, let $X_1^* = \Phi(X_1), Y_1^* = \Phi(Y_1), X^* = \Phi(X)$
We need to show the following four things:

(1) $P((Y_1 - X)(X_1 - X) > 0) - P((Y_1 - X)(-X_1 - X) > 0) =$
$2E\min(X_1^*, Y_1^*) - 2E\min(Y_1^*, 1 - X_1^*)$.

(2) $r_{mf} = \dfrac{\sum |n + 1 - p_i - i| - \sum |p_i - i|}{\left\lfloor \frac{n^2}{2} \right\rfloor}$ as n $\to \infty$ goes to

$2E|1 - \Phi(X_1) - \Phi(Y_1)| - 2E|\Phi(Y_1) - \Phi(X_1)| = 2E|1 - X_1^* - Y_1^*| - 2E|Y_1^* - X_1^*| =$
$2E\min(X_1^*, Y_1^*) - 2E\min(1 - X_1^*, Y_1^*)$.

(3) $2P((X_1 - X)(Y_1 - X) > 0) - 2P((Y_1 - X)(-X_1 - X) > 0) =$
$\dfrac{2}{p}\left[\sin^{-1}\left(\dfrac{1+r}{2}\right) - \sin^{-1}\left(\dfrac{1-r}{2}\right)\right]$.

(4) Combining (1), (2), and (3), the population value of the Modified Footrule Correlation coefficient, r_{mf}, is $\dfrac{2}{p}\left[\sin^{-1}\left(\dfrac{1+r}{2}\right) - \sin^{-1}\left(\dfrac{1-r}{2}\right)\right]$.

Part (1) case A, show that $P((Y_1 - X)(X_1 - X) > 0) = 2E\min(X_1^*, Y_1^*)$. The random variables X_1^*, Y_1^*, and X^* have U(0,1) distributions and that X^* is independent of X_1^* and Y_1^*. Also note that the quantities are ordinally invariant; i.e., $(Y_1 - X)(X_1 - X) > 0$ if and only if $(\Phi(Y_1) - \Phi(X))(\Phi(X_1) - \Phi(X)) = (Y_1^* - X^*)(X_1^* - X) > 0$. Now using these facts

(A)= $P((Y_1 - X)(X_1 - X) > 0) = P((Y_1^* - X^*)(X_1^* - X^*) > 0) =$
$E\left[P((Y_1^* - X^*)(X_1^* - X^*) > 0 | Y_1^*, X_1^*\right]$.

There are two ways that $(Y_1^* - X^*)(X_1^* - X^*) > 0$;

(a) $X^* < Y_1^*$, and $X^* < X_1^*$ or $X^* < \min(X_1^*, Y_1^*)$,

(b) $X^* > Y_1^*$, and $X^* > X_1^*$ or $X^* > \max(X_1^*, Y_1^*)$.

(a) occurs with conditional probability $\min(X_1^*, Y_1^*)$, and (b) with conditional probability $1 - \max(X_1^*, Y_1^*)$. Thus,

(A) = $E[\ \min(X_1^*, Y_1^*) + 1 - \max(X_1^*, Y_1^*)\ ]$, but $\min(X_1^*, Y_1^*) + \max(X_1^*, Y_1^*) = X_1^* + Y_1^*$ so that $E[\ \min(X_1^*, Y_1^*) + \max(X_1^*, Y_1^*)\ ] = E(X_1^*) + E(Y_1^*) = 1$. It follows that $E(\max(Y_1^*, X_1^*)) = 1 - E(\min(Y_1^*, X_1^*))$. Finally,

(A) = $E[\ \min(X_1^*, Y_1^*) + 1 - (1 - E(\min(Y_1^*, X_1^*))) = 2E(\min(Y_1^*, X_1^*))$.

Part (1) case B show $P((Y_1 - X)(-X_1 - X) > 0) = 2E(\min(Y_1^*, 1 - X_1^*))$.

(B) = $P((Y_1 - X)(-X_1 - X) > 0) = P((Y_1^* - X^*)(1 - X_1^* - X^*) > 0)$ since

$(Y_1 - X) > 0$ if and only if $\Phi(Y_1) - \Phi(X) = Y_1^* - X^* > 0$ and

$-X_1 - X > 0$ if and only if $\Phi(-X_1) - \Phi(X) = 1 - \Phi(X_1) - \Phi(X) = 1 - X_1^* - X^* > 0$,

using the symmetry of the normal distribution. Now conditioning on (X_1^*, Y_1^*)

(B) = $E\left[P((Y_1^* - X^*)(1 - X_1^* - X^*) > 0 \middle| Y_1^*, X_1^*\right]$. Now $(Y_1^* - X^*)(1 - X_1^* - X^*) > 0$ if $X^* < \min(Y_1^*, 1 - X_1^*)$ or if $X^* > \max(Y_1^*, 1 - X_1^*)$ which occur with probabilities $\min(Y_1^*, 1 - X_1^*)$ and $1 - \max(Y_1^*, 1 - X_1^*)$, respectively. So

$(B) = E\left[\min(Y_1^*, 1 - X_1^*) + 1 - \max(Y_1^*, 1 - X_1^*)\right] = 2E\left[\min(Y_1^*, 1 - X_1^*)\right]$. Combining parts A and B above the proof of (1) is complete;

$P((Y_1 - X)(X_1 - X) > 0) - P((Y_1 - X)(-X_1 - X) > 0) =$

$2E\min(X_1^*, Y_1^*) - 2E\min(Y_1^*, 1 - X_1^*)$.

Part (2): Let X and Y be a bivariate random variable for which a random sample has been ordered by the X variable $(x_{(i)}, y_{(p_i)}), i = 1, 2, \cdots, n$. When we go back to normal random variables, $X = X_1$ and $Y = Y_1$. Let $F_n(x) = \frac{\#x_i \le x}{n}$ be the empirical distribution function of X and similarly $G_n(y)$ for Y. Now $nF_n(x_{(i)}) = i$ and $nG_n(y_{(p_i)}) = p_i$. Asymptotically,

$\frac{\sum_{i=1}^{n} F_n(x_{(i)})}{n} = \frac{\sum(i/n)}{n}$ converges to $E(F(X)) = 1/2$, and $\left\lceil \frac{n^2}{2} \right\rceil$ is approximately $\frac{n^2}{2}$.

Thus, $\frac{\sum_{i=1}^{n} |p_i - i|}{\left\lceil \frac{n^2}{2} \right\rceil} \cong \frac{2}{n} \sum \left| \frac{p_i}{n} - \frac{i}{n} \right| = \frac{2}{n} \sum |G_n(y_{(p_i)}) - F_n(x_{(i)})|$ converges to

$2E|G(Y) - F(X)| = 2E|\Phi(Y_1) - \Phi(X_1)|$ with X_1 and Y_1 having correlation coefficient r.

Likewise, $\frac{\sum |n + 1 - p_i - i|}{\left\lceil \frac{n^2}{2} \right\rceil} \cong \frac{2}{n} \sum \left| \frac{n+1}{n} - G_n(y_{(p_i)}) - F_n(x_{(i)}) \right|$ converges to

$2E|1 - F(X) - G(Y)| = 2E|1 - \Phi(X_1) - \Phi(Y_1)|$. So, $\frac{\sum |n + 1 - p_i - i| - \sum |p_i - i|}{\left\lceil \frac{n^2}{2} \right\rceil}$ converges to $2E|1 - X_1^* - Y_1^*| - 2E|Y_1^* - X_1^*|$. But these last two terms can be rewritten as

$E|Y_1^* - X_1^*| = E\left[\max(Y_1^*, X_1^*) - \min(Y_1^*, X_1^*)\right] = E\left[1 - 2\min(Y_1^*, X_1^*)\right] =$

$1 - 2E\left[\min(Y_1^*, X_1^*)\right]$ and $E|1 - X_1^* - Y_1^*| = E\left[\max(Y_1^*, 1 - X_1^*) - \min(Y_1^*, 1 - X_1^*)\right] =$

$E\left[1 - 2\min(Y_1^*, 1 - X_1^*)\right] = 1 - 2E\left[\min(Y_1^*, 1 - X_1^*)\right]$. Finally,

$$\frac{\sum|n+1-p_i-i|-\sum|p_i-i|}{\left\lfloor \frac{n^2}{2} \right\rfloor} \text{ converges to } 2E|1-X_1^*-Y_1^*|-2E|Y_1^*-X_1^*| =$$

$1-2E[\min(Y_1^*,1-X_1^*)]$-($1-2E[\min(Y_1^*,X_1^*)]$)= $2E\min(Y_1^*,X_1^*)-2E\min(Y_1^*,1-X_1^*)$.

Part (3): In this part we modify the technique in Kruskal (1958) using the Quadrant correlation coefficient to compute probabilities. First we need to compute several quantities $\text{cov}(X_1-X,Y_1-X)=\text{cov}(X_1,Y_1)+\text{var}(X)=r+1$ and $\text{var}(X_1-X)=\text{var}(Y_1-X)=2$ so that $cor(X_1-X,Y_1-X)=\frac{1+r}{2}$. Similarly, $cor(-X_1-X,Y_1-X)=\frac{1-r}{2}$.

We now show how to compute the Quadrant correlation coefficient, Q, and then use this result. Let (W_1,W_2) be a bivariate random variable and let med = the median. Then Q for (W_1,W_2) is
$P((W_1-med(W_1))(W_2-med(W_2))>0)$ - $P((W_1-med(W_1))(W_2-med(W_2))<0)$.
Because the sum of these two probabilities is one, Q can be rewritten as
$2P((W_1-med(W_1))(W_2-med(W_2))>0)-1$.
For (W_1,W_2) as standardized normal random variables with correlation coefficient r, there is a result on page 290 of Cramer (1946) on the mass of the bivariate normal in the quadrants which shows that
$Q=2\left(\frac{2}{4}+\frac{2\sin^{-1}r}{2p}\right)-1=\frac{2\sin^{-1}r}{p}$.

We apply this to two sets of bivariate normal random variables: I; (Y_1-X,X_1-X) and II; $(Y_1-X,-X_1-X)$. The quadrant CC for I is $Q_I=2P((Y_1-X)(X_1-X)>0)-1=\frac{2}{p}\sin^{-1}(\frac{1+r}{2})$. The quadrant CC for II is $Q_{II}=2P((Y_1-X)(-X_1-X)>0)-1=\frac{2}{p}\sin^{-1}(\frac{1-r}{2})$. By subtracting these two results, we obtain the population value for the modified footrule CC; $Q_I-Q_{II}=2P((Y_1-X)(X_1-X)>0)-2P((Y_1-X)(-X_1-X)>0)=\frac{2}{p}(\sin^{-1}(\frac{1+r}{2})-\sin^{-1}(\frac{1-r}{2}))$.

It will now be shown that the inverse function is

$$r=\pm\tan(\frac{p}{4}r_{mf})\sqrt{1+2\cos(\frac{p}{2}r_{mf})}=\pm\sqrt{\frac{1-\cos\frac{p}{2}r_{mf}}{1+\cos\frac{p}{2}r_{mf}}(1+2\cos\frac{p}{2}r_{mf})}$$

proof: let $z=\frac{1+r}{2}$, then since, $z+\frac{1-r}{2}=\frac{1+r}{2}+\frac{1-r}{2}=1,\frac{1-r}{2}=1-z$. It follows that

$\frac{p}{2} r_{mf} = \sin^{-1} z - \sin^{-1}(1-z)$. Now let $\frac{p}{2} r_{mf} = w$ and take the cosine of both sides of the equation

$\cos w = \cos(\sin^{-1} z - \sin^{-1}(1-z)) =$

$\cos(\sin^{-1} z)\cos(\sin^{-1}(1-z)) + \sin(\sin^{-1} z)\sin(\sin^{-1}(1-z))$

$= \sqrt{1-z^2}\sqrt{1-(1-z^2)} + z(1-z)$. Now isolate the square root term on one side of the equation and square and collect terms. The result is the quadratic equation in z,

$z^2 - z + \frac{\cos^2 w}{2(1+\cos w)} = 0$; this holds except at $w = \frac{-p}{2}$ or $r_{mf} = -1$ since then

$\cos(\frac{-p}{2}) = -1$. By the symmetry of the correlation function if it is negative, the inverse can be found for the positive number and then the negative taken. Now the quadratic equation is solved,

$z = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - \frac{2\cos^2 w}{1+\cos w}} = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{1-\cos w}{1+\cos w}(1+2\cos w)}$. Since $z = \frac{1+r}{2}$,

$\frac{1+r}{2} = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{1-\cos w}{1+\cos w}(1+2\cos w)}$ or

$r = \pm\frac{1}{2}\sqrt{\frac{1-\cos w}{1+\cos w}(1+2\cos w)} = \pm\tan\frac{w}{2}\sqrt{1+\cos w}$. Now substituting $\frac{p}{2} r_{mf} = w$, the result follows.

The Population Value of r_{av} for the Bivariate Normal Distribution

$r_{av} = \frac{\sum\left|\frac{x_i - \bar{x}}{SA_x} + \frac{y_i - \bar{y}}{SA_y}\right| - \sum\left|\frac{y_i - \bar{y}}{SA_y} - \frac{x_i - \bar{x}}{SA_x}\right|}{2}$ and the asymptotic or population value of this correlation coefficient is

$r_{av} = \frac{\sqrt{1+r} - \sqrt{1-r}}{\sqrt{2}}$.

mmProof: For the bivariate normal random variable (X,Y) with means m_x, m_y, variances s_x^2, s_y^2, and correlation r, it is a straight-forward integration to show that $E|X - m_x| = \sqrt{\frac{2}{p}}\, s_x$ and similarly for Y. Just as the population value of Pearson's r is defined by replacing summations by expectations, so is the population value of r_{av} defined. The term $\frac{SA_x}{n} = \frac{\sum|x_i - \bar{x}|}{n}$ is converging to $E|X - m_x|$ and a similar result hold for $\frac{SA_y}{n}$.

The term $\sum\left|\frac{x_i-\bar{x}}{SA_x}+\frac{y_i-\bar{y}}{SA_y}\right|$ becomes in the limit $E\left|\frac{X-m_x}{s_x\sqrt{\frac{2}{p}}}+\frac{Y-m_y}{s_y\sqrt{\frac{2}{p}}}\right|$. Let $Z_1=\frac{X-m_x}{s_x}$ and $Z_2=\frac{Y-m_y}{s_y}$. Then the population value of r_{av} is $r_{av}=\frac{1}{2}\sqrt{\frac{p}{2}}\left\{E|Z_1+Z_2|-E|Z_2-Z_1|\right\}$

Since Z_1 and Z_2 are standarized normal random variables with correlation coefficient r, Z_1+Z_2 has a $N(0,2(1+r))$ distribution and Z_1-Z_2 has a $N(0,2(1-r))$ distribution. Now $E|Z_1+Z_2|=\sqrt{\frac{2}{p}}\sqrt{2(1+r)}=2\sqrt{\frac{1+r}{p}}$ and $E|Z_1-Z_2|=\sqrt{\frac{2}{p}}\sqrt{2(1-r)}=2\sqrt{\frac{1-r}{p}}$. Thus,

$r_{av}=\frac{1}{2}\sqrt{\frac{p}{2}}\frac{2}{\sqrt{p}}\left(\sqrt{1+r}-\sqrt{1-r}\right)=\frac{\sqrt{1+r}-\sqrt{1-r}}{\sqrt{2}}$.

THE ASYMPTOTIC DISTRIBUTION OF r_{av} FOR THE BIVARIATE NORMAL

First, some facts necessary for the result are collected. Let X,Y be defined as in the previous section and for random sample $\{X_i,Y_i\}_{i=1}^n$ let $Z_i=\frac{X_i-m_x}{s_x}$ and $W_i=\frac{Y_i-m_y}{s_y}$, $i=1,2,\cdots,n$. Then $E(Z_i)=E(W_i)=0, V(Z_i)=V(W_i)=1$, $\text{cov}(Z_i,W_i)=r$, and it is easily shown that $\text{cov}(Z_i+W_i,W_j-Z_j)=0$ for i=j and also for $i\neq j$. Thus, Z_i+W_i and W_i-Z_i are independent, and hence, so are $|Z_i+W_i|$ and $|W_i-Z_i|$. It follows that $\text{cov}(|Z_i+W_i|,|W_j-Z_j|)=0\forall i,j$. The variances are needed for the absolute values of these two random variables; $Z_i+W_i\sim N(0,2(1+r))$ and $W_i-Z_i\sim N(0,2(1-r))$.

$V(|Z_i+W_i|)=E(Z_i+W_i)^2-E^2|Z_i+W_i|=V(Z_i+W_i)-E^2|Z_i+W_i|=2(1+r)-4\left(\frac{1+r}{p}\right)$

$=\frac{2(1+r)(p-2)}{p}$.

$V(|Z_i-W_i|)=E(Z_i-W_i)^2-E^2|Z_i-W_i|=V(Z_i-W_i)-E^2|Z_i-W_i|=2(1-r)-4\left(\frac{1-r}{p}\right)$

$=\frac{2(1-r)(p-2)}{p}$.

Because the covariance is zero, the variance of the sum $|Z_i+W_i|-|W_i-Z_i|$ is the sum of the above variances which is $\frac{4(p-2)}{p}$.

Theorem: The asymptotic distribution of $\sqrt{n}r_{av}$ is $N(r_{av},\frac{p-2}{2})$.

In what follows the Slutsky theorems concerning the relationship between convergence in probability and convergence in distribution are used, see page 254 of Cramer (9th printing).

Write the absolute value CC as follows:

$$r_{av} = \frac{1}{2}\left\{\sum\left|\frac{x_i-\bar{x}}{nSA_x/n}+\frac{y_i-\bar{y}}{nSA_y/n}\right|-\sum\left|\frac{y_i-\bar{y}}{nSA_y/n}-\frac{x_i-\bar{x}}{nSA_x/n}\right|\right\}.$$

Now $\frac{SA_x}{n}\xrightarrow{p} s_x\sqrt{2/p}, \frac{SA_y}{n}\xrightarrow{p} s_y\sqrt{2/p}, \bar{x}\xrightarrow{p} m_x, \bar{y}\xrightarrow{p} m_y$. Thus, r_{av} is converging in probability to $\frac{1}{2n}\sqrt{\frac{p}{2}}\left\{\sum\left|\frac{x_i-\bar{x}}{s_x}+\frac{y_i-\bar{y}}{s_y}\right|-\sum\left|\frac{y_i-\bar{y}}{s_y}-\frac{x_i-\bar{x}}{s_x}\right|\right\}$, and in terms of random variables this is $\frac{1}{2n}\sqrt{\frac{p}{2}}\{\sum|Z_i+W_i|-\sum|W_i-Z_i|\}$. The two inner summations can be written as $S_n = \sum(|Z_i+W_i|-|W_i-Z_i|)$. By the central limit theorem, this sum of independent random variables with finite mean, m, and variance, s^2, has a limiting normal distribution.

Now $m = E(|Z_i+W_i|-|W_i-Z_i|) = 2\sqrt{\frac{1+r}{p}}-2\sqrt{\frac{1-r}{p}}$ and

$s^2 = V(|Z_i+W_i|-|W_i-Z_i|) = \frac{4(p-2)}{p}$. Then the central limit theorem applied to S_n gives $\frac{S_n-nm}{\sqrt{n}s}$ has an approximate N(0,1) distribution or $\frac{S_n}{\sqrt{n}}$ has an approximate $N(\sqrt{n}m, s^2)$. Now $\sqrt{n}r_{av}\xrightarrow{p}\frac{1}{2}\sqrt{\frac{p}{2}}\frac{S_n}{\sqrt{n}}$ and because

$E(\sqrt{n}r_{av})\xrightarrow{p}\frac{1}{2}\sqrt{\frac{p}{2}}\sqrt{n}2(\sqrt{\frac{1+r}{p}}-\sqrt{\frac{1-r}{p}}) = \sqrt{n}\left(\frac{\sqrt{1+r}-\sqrt{1-r}}{\sqrt{2}}\right) = \sqrt{n}r_{av}$, and

$V(\sqrt{n}r_{av})\xrightarrow{p}\frac{ps^2}{8} = \frac{p-2}{2}$, the limiting distribution of $\sqrt{n}(r_{av}-r_{av})$ is $N(0,\frac{p-2}{2})$.

A Simple comparison of Some Correlation Coefficients

For the bivariate normal distribution a table is given to compare the population values of several correlation coefficients. In addition to the correlations above, Spearmans r_s and Kendall's r_k which appear in Kruskal (1958) as well as the Greatest Deviation r_{gd} Gideon and Hollister (1987) are also listed. For Spearman, $r_s = \frac{6}{p}\sin^{-1}(r/2)$; for Kendall and the Greatest Deviation, $r_k = r_{gd} = \frac{2}{p}\sin^{-1}(r)$; Pearson's is just r. The quadrant correlation coefficient is also the same as Kendall's.

Population values of some correlation coefficients for the bivariate normal

r	0	.25	.50	.75	.90	.95	.99	.999	1
s	0	.2394	.4826	.7341	.8915	.9453	.9890	.9989	1
av	0	.1782	.3660	.5819	.7511	.8293	.9268	.9774	1
mf	0	.1851	.3790	.5985	.7660	.8414	.9331	.9795	1
k, gd	0	.1609	.3333	.5399	.7129	.7978	.9099	.9715	1

In order to see how close the sample values of r_{mf} are to its population values, 100 simulations of it were made on each setting of r_{mf} and a sample size. The average value appears in the table below. It is clear that r_{mf} is biased on the low side of r_{mf} when $r_{mf}>0$ and that the bias is already quite small for sample sizes of 20 or more.

		average of 100 simulations of r_{mf}					
	r	0	.1	.25	.50	.75	.95
	r_{mf}	0	.0736	.1851	.3790	.5985	.8414
n=10	r_{mf}	.0084	.0724	.1812	.3416	.5460	.8124
n=20	r_{mf}	-.0183	.0657	.1887	.3668	.5892	.8236
n=50	r_{mf}	-.0030	.0778	.1924	.3857	.5924	.8370

ASYMPTOTIC RELATIONSHIP BETWEEN A CORRELATION COEFFICIENT AND ITS ASSOCIATED SLOPE ESTIMATE IN A SIMPLE LINEAR REGRESSION

Let $r(x,y)$ be a correlation coefficient on data (x,y) and $r(X,Y)$ the corresponding population CC. Assume the simple linear regression model: $E(Y|X=x)=m_2+b(H)(x-m_1)$ where $b(H)$ is the theoretical slope with H being the bivariate distribution of (X,Y). Abbreviate $r(X,Y-Xb)$ to r_b and expand it into a truncated Taylor series about $b(H)$**;**

$r(X,Y-Xb)\cong r(X,Y-Xb(H))+\frac{d}{db}r(X,Y-Xb)\Big|_{b=b(H)}(b-b(H))$ or letting

$r'_{b(H)}=\frac{d}{db}r(X,Y-Xb)\Big|_{b=b(H)}$ this can be shorten to $r_b\cong r_{b(H)}+r'_{b(H)}(b-b(H))$. Now for sample size n, let $r(x,y-x\hat{b})=0$ so that $\hat{b}$ is the slope estimate of $b(H)$ for CC r. Assume it is known that $\sqrt{n}r(w_1,w_2)$ for w_1 and w_2 independent is asymptotically $N(0,s_r^2)$**.** Also assume that for H, X and Y-X$b(H)$ are independent random variables. Then in the truncated Taylor series above, for large n, replace r by r and b by $\hat{b}$ and obtain $r(x,y-x\hat{b})=0\cong r(x,y-xb(H))+r'_{b(H)}(\hat{b}-b(H))$**.** Thus, approximately, $(-1)\sqrt{n}r'_{b(H)}(\hat{b}-b(H))=\sqrt{n}r(x,y-xb(H))\overset{d}{=}N(0,s_r^2)$ and the approximate large sample distribution of $\hat{b}$ is $N\left(b(H),\frac{s_r^2}{n(r'_{b(H)})^2}\right)$**.** This result is now applied to most of the CC that have been defined.

Let r be Pearson's CC and H a bivariate normal distribution with parameters, m_1,m_2,s_1,s_2,r where subscript 1 is for X and 2 for Y. Then $b(H)=r\frac{s_2}{s_1}$**,** $r(X,Y-Xb(H))=0$, and $\sqrt{n}r(x,y-xb(H))$ is asymptotically N(0,1),(Anderson, 1958, theorem 4.2.6 page 77). Also, $r_b\equiv r(X,Y-bX)=\frac{rs_2-bs_1}{s_{res}}$ where

$s_{res} = (s_2^2 - 2b\,r\,s_1 s_2 + b^2 s_1^2)^{1/2}$. Note that $r_{b(H)} = 0$, $r'_b = \dfrac{-s_1 s_2^2(1-r^2)}{s_{res}^3}$, and $r'_{b(H)} = \dfrac{-s_1}{s_2\sqrt{1-r^2}} < 0$. These results are now related to the general results given earlier for the r estimate (i.e. least squares) of slope. The variance $s_r^2 = 1$, $(r'_{b(H)})^2 = \dfrac{s_1^2}{s_2^2(1-r^2)}$, and so the asymptotic distribution of $\hat{b}$ is $N\left(b(H), \dfrac{s_2^2(1-r^2)}{n s_1^2}\right)$; for $s_1^2 = s_2^2 = 1$, the variance becomes $\dfrac{1-r^2}{n}$. To connect this to the usual fixed x regression, let the residual variance be $s^2 = s_2^2(1-r^2)$ and $n s_1^2 \cong \sum (x_i - \bar{x})^2$ so that in this form the approximate large sample distribution of $\hat{b}$ is $N\left(b(H), \dfrac{s^2}{\sum (x_i - \bar{x})^2}\right)$.

The procedure is now carried out for r_{gd}, the Greatest Deviation CC for the standardized bivariate normal distribution. In this case $b(H) = r$. To obtain the asymptotic distribution of the r_{gd} slope estimate, the following quantities are needed: $r_{gd,b} = r_{gd}(X, Y - bX) = \dfrac{2\sin^{-1} r_b}{p}$ and $\dfrac{d}{db} r_{gd,b} = \dfrac{2r'_b}{p\sqrt{1-r_b^2}}$ so that $r'_{gd,b(H)} = \dfrac{-2}{p\sqrt{1-r^2}}$. From Gideon et al. (1989), $\sqrt{n}\, r_{gd}(x, y - rx)$ is asymptotically N(0,1) so that $s_{gd}^2 = 1$. Finally, $\hat{b}_{gd}$ is approximately $N\left(r, \dfrac{p^2(1-r^2)}{4n}\right)$.

For Kendall's tau which is denoted here by r_k, $r_{k,b} = r_{gd,b} = \dfrac{2\sin^{-1} r_b}{p}$. Thus, the derivative term is the same as for the Greatest Deviation CC, $r'_{b(H)} = \dfrac{-2}{p\sqrt{1-r^2}}$. However, $\sqrt{n}\, r_k(x, y - x\,b(H))$ is asymptotically N(0,4/9) so that $s_k = 2/3$. Then the variance term is $\dfrac{s_k^2}{n(r'_{b(H)})^2} = \dfrac{4p^2(1-r^2)}{4 * 9n}$ so that the asymptotic of $\hat{b}_k$ is $N\left(r, \dfrac{p^2(1-r^2)}{9n}\right)$.

For Spearman's rho, r_S, the following facts hold: $r_s = \dfrac{6\sin^{-1}(r/2)}{p}$ so that $r_{s,b} = \dfrac{6\sin^{-1}(r_b/2)}{p}$ and $r'_{s,b} = \dfrac{6}{p\sqrt{1-(r_b^2/2)}} * \dfrac{r'_b}{2}$. Thus, $r'_{s,b(H)} = \dfrac{-3}{p\sqrt{1-r^2}}$. It is

known that the asymptotic null distribution of $\sqrt{n-1}r_s$ is N(0,1) and so $s_s = 1$. Finally, the asymptotic distribution of $\hat{b}_s$ is $N\left(b(H), \frac{p^2(1-r^2)}{9n}\right)$.

For the modified footrule CC, rmf, $r_{mf,b} = \frac{2}{p}\left[\sin^{-1}\left(\frac{1+r_b}{2}\right) - \sin^{-1}\left(\frac{1-r_b}{2}\right)\right]$ and $r'_{mf,b} = \frac{2}{p}\left[\frac{r'_b}{2\sqrt{1-((1+r_b)/2)^2}} - \frac{-r'_b}{2\sqrt{1-((1-r_b)/2)^2}}\right]$. This quantity is now evaluated at $b(H)$ to obtain $r'_{mf,b(H)} = \frac{2}{p}\left[\frac{(-1/\sqrt{1-r^2})}{2\sqrt{1-((1+0)/2)^2}} - \frac{(-1/\sqrt{1-r^2})}{2\sqrt{1-((1-0)/2)^2}}\right] = \frac{-4}{p\sqrt{3}\sqrt{1-r^2}}$. The asymptotic null distribution of $\sqrt{n}r_{mf}$ is N(0,2/3) so that $s^2_{mf} = 2/3$. Because $\frac{s^2_{mf}}{n(r'_{mf,b(H)})^2} = \frac{p^2(1-r^2)}{8n}$ the asymptotic distribution of $\hat{b}_{mf}$ is $N\left(b(H), \frac{p^2(1-r^2)}{8n}\right)$.

For the Absolute deviation CC let X,Y independent r.v.'s, then the limiting distribution of $\sqrt{n}r_{av}$ is $N(0, \frac{p-2}{2})$ so that $s^2_{av} = \frac{p-2}{2}$. The following are needed for the asymptotic distribution of the slope estimate $\hat{b}_{av}$.
$r_{av,b} = \frac{\sqrt{1+r_b} - \sqrt{1-r_b}}{\sqrt{2}}$ and its derivative $r'_{av,b} = \frac{r'_b}{\sqrt{2(1+r_b)}}$. The derivative evaluated at the population value $b(H)$ is $r'_{av,b(H)} = \frac{r'_{b(H)}}{\sqrt{2}} = \frac{-s_1}{s_2\sqrt{1-r^2}\sqrt{2}}$. So for the asymptotic variance of $\hat{b}_{av}$, we obtain $\frac{s^2_{av}}{n(r'_{av,b(H)})^2} = \frac{s_2^2(1-r^2)(p-2)}{ns_1^2}$. The ratio of the asymptotic SD of $\hat{b}_{av}$ to $\hat{b}_r$ is, since $V(\hat{b}_{av}) = (p-2)V(\hat{b}_r)$, $\sqrt{p-2} = \sqrt{1.1416} = 1.0685$.

A table to compare the ratios of the asymptotic standard deviations of the slope estimators for each of the CC to Pearson's CC (i.e. to $\sqrt{1-r^2}/\sqrt{n}$) is now given.

CC:	Greatest Deviation	Spearman	Modified Footrule	Kendall	Absolute Deviation
ratio:	$p/2 = 1.5708$	$p/3 = 1.047$	$p/(2\sqrt{2}) = 1.1107$	$p/3 = 1.047$	$\sqrt{p-2}$ = 1.0685

REGRESSION AND CORRELATION DUALITY

In this section it is shown that the following maxima and minima with respect to b are equivalent: min(res), max$(r'_b)^2$, and min(var($\hat{b}$)). The general bivariate normal

distribution is considered for rgd. From the preceding section, $r_b = \frac{r s_2 - b s_1}{s_{res}}$, $\sqrt{1-r_b^2} = \frac{s_2\sqrt{1-r^2}}{s_{res}}$, and $r_b' = \frac{-s_1 s_2^2(1-r^2)}{s_{res}^3}$. From these it follows that $r_{gd,b}' = \frac{2r_b'}{p\sqrt{1-r_b^2}} = \frac{-2s_1 s_2\sqrt{1-r^2}}{p s_{res}^2}$, and at $b(H) = \frac{r s_2}{s_1}$, $s_{res}^2 = s_2^2(1-r^2)$. So for CC rgd, $r_{gd,b(H)}' = \frac{-2s_1}{p s_2\sqrt{1-r^2}}$ so that the asymptotic distribution of $\hat{b}_{gd}$ is $N\left(\frac{r s_2}{s_1}, \frac{p^2 s_2^2(1-r^2)}{4n s_1^2}\right)$.

Note that $\text{var}(\hat{b}_{gd}) = \frac{p^2}{4}\text{var}(\hat{b}_r)$ since $\text{var}(\hat{b}_r) = \frac{s_2^2(1-r^2)}{s_1^2}$. In a similar manner $\text{var}(\hat{b}_k) = \text{var}(\hat{b}_s) = \frac{p^2}{9}\text{var}(\hat{b}_r)$, $\text{var}(\hat{b}_{mf}) = \frac{p^2}{8}\text{var}(\hat{b}_r)$, and $\text{var}(\hat{b}_{av}) = (p-2)\text{var}(\hat{b}_r)\,\text{var}(\hat{b}_{av}) = (p-2)\text{var}(\hat{b}_r)$.

Now s_{res}^2 is minimized at $b = \frac{r s_2}{s_1}$ and because $(r_b')^2 = \frac{s_1^2 s_2^4(1-r^2)^4}{s_{res}^6}$, it is maximized, and hence, $\text{var}(\hat{b}_{cc})$ which is proportional to the reciprocal of $(r_b')^2$ is minimized. The cc subscript on $\hat{b}$ means any of the above CC could be used.