Statistics of Range of a Set of Normally Distributed Numbers

Charles R. Schwarz[1]

Abstract: We consider a set of numbers independently drawn from a normal distribution. We investigate the statistical properties of the maximum, minimum, and range of this set. We find that the range is closely related to the standard deviation of the original population. In particular, we investigate the use of the online positioning user service (OPUS) global positioning system (GPS) precise position utility, which produces three estimates of each coordinate and reports the range of these three estimates. We find that the range divided by 1.6926 is an unbiased estimate of the standard deviation of a single coordinate estimate, and that the variance of this estimate is 0.2755 σ^2. We compare this to the more conventional method of estimating the standard deviation of a single observation from the sum of squares of residuals, which is shown to have a variance 0.2275 σ^2.

DOI: 10.1061/(ASCE)0733-9453(2006)132:4(155)

CE Database subject headings: Statistics; surveys.

Introduction

This investigation was motivated by users of the online positioning user service (OPUS) utility for computing precise coordinates for global positioning system (GPS) geodetic receivers (see http://www.ngs.noaa.gov/OPUS). This utility, provided by the U.S. National Geodetic Survey, computes highly accurate GPS positions from a single data set provided by the user. It performs a differential GPS solution by searching for suitable reference station data in the archive of Continuously Operating Reference Station (CORS) data. OPUS selects three CORS stations and performs three single baseline solutions. This produces three estimates for the coordinates of the user's receiver in the U.S. National Spatial Reference System. The OPUS utility reports the mean of these three estimates, together with their range, to the user. Fig. 1 shows a typical solution report obtained from OPUS. The numbers following the coordinates show the range of the three estimates of each coordinate.

Each single baseline solution is highly accurate by itself. The reason to use three such baselines appears to be to provide some protection against blunders, not because they provide significant statistical or geometric redundancy. The range of the three estimates tells the user if blunders are present. If the range is significantly larger than normally obtained with similar data sets, then one or more of the reference station's coordinates or tracking data may have contained a blunder.

There are several ways to estimate the uncertainty of the mean of the three single baseline estimates reported to the user. Among these are

1. Linear error propagation, using the variances of the individual single baseline solutions and assuming that these estimates are statistically independent. If x_1, x_2, and x_3 are three independent estimates of coordinate X, with variances σ_1^2, σ_2^2, and σ_3^2, then the best estimate of the coordinate X is the weighted mean

$$\bar{x} = \frac{\sum_{i=1}^{3} w_i x_i}{\sum_{i=1}^{3} w_i}$$

where the weights are $w_i = 1/\sigma_i^2$. The variance of the mean is then

$$\sigma_{\bar{x}}^2 = \frac{1}{\sum_{i=1}^{3} w_i}$$

According to the OPUS user documentation, the designers of OPUS declined to use this statistic, saying that the variances from the single baseline solutions are notoriously overoptimistic (http://www.ngs.noaa.gov/OPUS/Using_OPUS.html#accuracy).

2. Linear error propagation, using external estimates of the variances of single baseline solutions. A common method of obtaining such external estimates is to perform a study using a large number of single baseline solutions. Such studies may produce rules for estimating variances from some property of the data, such as the time span of the data or the length of the baseline.
3. Linear error propagation, assuming that all three single baseline solutions have the same variance σ^2. Then the best estimate of the coordinate X is the simple mean

$$\bar{x} = \frac{\sum_{i=1}^{3} x_i}{3}$$

and its variance is

[1]Consultant, Geodesy, 5320 Wehawken Rd., Bethesda, MD 20816. E-mail: charlies2@earthlink.net

Note. Discussion open until April 1, 2007. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on May 10, 2005; approved on July 8, 2005. This paper is part of the ***Journal of Surveying Engineering***, Vol. 132, No. 4, November 1, 2006. ©ASCE, ISSN 0733-9453/2006/4-155–159/$25.00.

```
      USER: charlies2@earthlink.net                    DATE: April 29, 2005
RINEX FILE: bkly118o.05o                               TIME: 12:09:34 UTC

  SOFTWARE: page5  0411.19 master24.pl                START: 2005/04/28  14:00:00
 EPHEMERIS: igu13204.eph [ultra-rapid]                 STOP: 2005/04/28  15:00:00
  NAV FILE: brdc1180.05n                           OBS USED:  2120 /  2173   :  98%
  ANT NAME: NONE                                 # FIXED AMB:    18 /    18   : 100%
ARP HEIGHT: 0.0                                  OVERALL RMS: 0.010(m)

 REF FRAME: NAD_83(CORS96)(EPOCH:2002.0000)             ITRF00 (EPOCH:2005.3222)

         X:       776242.978(m)   0.002(m)          776242.306(m)   0.002(m)
         Y:     -4986708.281(m)   0.026(m)        -4986706.835(m)   0.026(m)
         Z:      3888159.007(m)   0.015(m)         3888158.876(m)   0.015(m)

       LAT:   37 47 52.09045      0.011(m)        37 47 52.11755      0.011(m)
     E LON:  278 50 52.04960      0.005(m)       278 50 52.03155      0.005(m)
     W LON:   81  9  7.95040      0.005(m)        81  9  7.96845      0.005(m)
    EL HGT:           692.689(m)   0.026(m)               691.398(m)   0.026(m)
 ORTHO HGT:           723.765(m)   0.036(m) [Geoid03 NAVD88]

                         UTM COORDINATES    STATE PLANE COORDINATES
                          UTM (Zone 17)        SPC (4702 WV S)
Northing (Y)  [meters]     4183392.127          88557.655
Easting (X)   [meters]      486599.893         586595.234
Convergence   [degrees]    -0.09328516         -0.09409456
Point Scale                 0.99960221          0.99994831
Combined Factor             0.99949357          0.99983964

US NATIONAL GRID DESIGNATOR: 17SMB8660083392(NAD 83)


                              BASE STATIONS USED
PID       DESIGNATION                         LATITUDE    LONGITUDE DISTANCE(m)
DF4048 GALP GALLIPOLIS CORS ARP              N385039.148 W0821640.092  152255.7
DF5767 DOBS DOBSON CORS ARP                  N362531.514 W0804311.711  157086.8
AI1571 BLKV BLACKSBURG CORS ARP              N371221.637 W0802452.276   92575.9
```

Fig. 1. Portion of an OPUS data sheet

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{3}$$

While linear error propagation is formally correct, it is easily contaminated by blunders in the data. Detecting such blunders appears to have been the main concern of the designers of OPUS. Their methodology is to estimate the coordinate X with the simple mean of the three single baseline solutions, but report the range of the three estimates rather than a standard deviation. The range (also called the peak-to-peak error) will show the presence of a blunder more clearly than a propagated variance. The explanation on the OPUS web site suggests that the peak-to-peak error is intended to be a rough guide to the accuracy of the reported coordinates.

A number of OPUS users have asked for a formal standard deviation or variance for the reported coordinates. This has prompted interest in the question of whether such a standard deviation can be computed from the reported range of the three single baseline solutions. Intuitively, it seems that such a relationship must exist; if the uncertainties of the single baseline solutions are large, then one would expect the range of a sample of three of them to be large.

Statistics of the Maximum, Minimum, and Range

Let $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ be three numbers independently drawn from a population with a normal distribution with mean μ and standard deviation σ; i.e., $\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3 \sim n(\mu,\sigma)$. Standardize these by $\mathbf{x}_1=(\mathbf{X}_1-\mu)/\sigma$, $\mathbf{x}_2=(\mathbf{X}_2-\mu)/\sigma$, and $\mathbf{x}_3=(\mathbf{X}_3-\mu)/\sigma$ so that $\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3 \sim n(0,1)$.

Let $\mathbf{u}=\max(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)$. Then $\mathbf{u}$ is also a random variable. Its distribution is the extreme value distribution, a topic treated in the subject of order statistics (see Wilks 1962). Its expected value [see Weisstein, undated, Eq. (34)] is

$$E[\mathbf{u}] = \int_{-\infty}^{\infty} u f_{\mathbf{u}}(u)du = \frac{3}{2\sqrt{\pi}} = 0.8463$$

and its variance [Weisstein, undated, Eq. (39)] is

$$E[(\mathbf{u}-E[\mathbf{u}])^2] = \int_{-\infty}^{\infty} u^2 f_{\mathbf{u}}(u)du - E^2[\mathbf{u}] = \frac{4\pi - 9 + 2\sqrt{3}}{4\pi} = 0.5593$$

Similarly, let $\mathbf{v}=\min(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)$. It is easy to show that

$$E[\mathbf{v}] = -E[\mathbf{u}]$$

Finally, let $\mathbf{w}$ be the range of the standardized random variables. Then

$$E[\mathbf{w}] = E[\mathbf{u}-\mathbf{v}] = E[\mathbf{u}] - E[\mathbf{v}] = 2E[\mathbf{u}]$$

Application to the OPUS Utility

Returning to the original set of numbers $\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3$ and computing their maximum, minimum, and range, we have

$$E[\mathbf{max}] = \mu + \frac{3}{2\sqrt{\pi}}\sigma$$

$$E[\mathbf{min}] = \mu - \frac{3}{2\sqrt{\pi}}\sigma \text{ and}$$

$$E[\mathbf{range}] = \frac{3}{\sqrt{\pi}}\sigma = 1.6926\sigma$$

This means that

$$\mathbf{s}_1 = \mathbf{range}/1.6929$$

is an unbiased estimate of σ, the standard deviation of the population from which the three numbers were drawn. Furthermore,

$$\mathbf{s}_1/\sqrt{3} = \mathbf{range}/2.9317$$

may be used as an unbiased estimate of the standard deviation of the mean of the three individual estimates.

How Good Is This Estimate?

If we want to use **range**/1.6926 as an estimate of σ, it makes sense to ask how good is this estimate.

We have already noted that

$$\mathrm{Var}(\mathbf{u}) = E[(\mathbf{u}-E[\mathbf{u}])^2] = E[\mathbf{u}^2] - (E[\mathbf{u}])^2 = 0.5593$$

We can also easily show that $\mathrm{Var}(\mathbf{v})=\mathrm{Var}(\mathbf{u})$.

With $\mathbf{w}=\mathbf{u}-\mathbf{v}$ as the range of the standardized random variables we have

$$\mathrm{Var}[\mathbf{w}] = E[(\mathbf{w}-E[\mathbf{w}])^2] = \mathrm{Var}[\mathbf{u}] + \mathrm{Var}[\mathbf{v}] - 2\mathrm{Cov}[\mathbf{u},\mathbf{v}]$$

We are tempted to assume that the covariance between the maximum $\mathbf{u}$ and minimum $\mathbf{v}$ should be zero. After all, the maximum is one of the random variables $\mathbf{x}_i$ and the minimum is another one. $\mathbf{x}_j$, and $\mathbf{x}_i$ and $\mathbf{x}_j$ are statistically independent. However, the analysis in the Appendix shows that this is not the case. Instead, the covariance is

$$\mathrm{Cov}[\mathbf{u},\mathbf{v}] = E[(\mathbf{u}-E[\mathbf{u}])(\mathbf{v}-E[\mathbf{v}])] = E[\mathbf{uv}] - E[\mathbf{u}]E[\mathbf{v}]$$

where

$$E[\mathbf{uv}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv f_{uv}(u,v)\,du\,dv$$

Using the expression for $f_{uv}(u,v)$ from the Appendix and setting $n=3$

$$E[\mathbf{uv}] = 6\int_{-\infty}^{\infty}\int_{v}^{\infty}[F_{\mathbf{x}}(u)-F_{\mathbf{x}}(v)]f_{\mathbf{x}}(u)f_{\mathbf{x}}(v)\,du\,dv$$

For the normal distribution, the probability density function is the Gaussian function

$$f_{\mathbf{x}}(u) = \mathrm{G}(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$$

and the distribution function is

$$F_{\mathbf{x}}(u) = \frac{1}{2} + \frac{1}{2}\mathrm{Erf}(u/\sqrt{2})$$

where $\mathrm{Erf}(z)=2/\sqrt{\pi}\int_0^z e^{-t^2}dt$ is the error function described in many texts (see, e.g., Abramowitz and Stegun 1977). Thus,

$$\begin{aligned}
E[\mathbf{uv}] &= 3\int_{-\infty}^{\infty}\int_{v}^{\infty} uv[\mathrm{Erf}(u/\sqrt{2}) - \mathrm{Erf}(v/\sqrt{2})]\mathrm{G}(u)\mathrm{G}(v)\,du\,dv \\
&= 3\int_{-\infty}^{\infty} v\mathrm{G}(v)\left[\int_{v}^{\infty} u\mathrm{Erf}(u/\sqrt{2})\mathrm{G}(u)\,du\right]dv \\
&\quad - 3\int_{-\infty}^{\infty} v\mathrm{G}(v)\mathrm{Erf}(v/\sqrt{2})\left[\int_{v}^{\infty} u\mathrm{G}(u)\,du\right]dv \\
&= 3\int_{-\infty}^{\infty} u\mathrm{G}(u)\mathrm{Erf}(u/\sqrt{2})\left[\int_{-\infty}^{u} v\mathrm{G}(v)\,dv\right]du \\
&\quad - 3\int_{-\infty}^{\infty} v\mathrm{G}(v)\mathrm{Erf}(v/\sqrt{2})\left[\int_{v}^{\infty} u\mathrm{G}(u)\,du\right]dv \\
&= 3\int_{-\infty}^{\infty} u\mathrm{G}(u)\mathrm{Erf}(u/\sqrt{2})[-\mathrm{G}(u)]\,du \\
&\quad - 3\int_{-\infty}^{\infty} v\mathrm{G}(v)\mathrm{Erf}(v/\sqrt{2})[\mathrm{G}(v)]\,dv \\
&= -6\int_{-\infty}^{\infty} z\mathrm{G}^2(z)\mathrm{Erf}(z/\sqrt{2})\,dz \\
&= -\frac{3}{\pi}\int_{-\infty}^{\infty} ze^{-z^2}\mathrm{Erf}(z/\sqrt{2})\,dz = -\frac{3}{\pi}\frac{1}{\sqrt{3}} = -\frac{\sqrt{3}}{\pi} = -0.5513
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{Cov}[\mathbf{uv}] &= E[\mathbf{uv}] - E[\mathbf{u}]E[\mathbf{v}] \\
&= -\frac{\sqrt{3}}{\pi} + \left(\frac{3}{2\sqrt{\pi}}\right)^2 = \frac{9-4\sqrt{3}}{4\pi} = 0.1649
\end{aligned}$$

and the variance of the range of the standardized random variables is

$$\begin{aligned}
\mathrm{Var}[\mathbf{w}] &= \mathrm{Var}[\mathbf{u}] + \mathrm{Var}[\mathbf{v}] - 2\mathrm{Cov}[\mathbf{u},\mathbf{v}] \\
&= 2\frac{4\pi - 9 + 2\sqrt{3}}{4\pi} - 2\frac{9-4\sqrt{3}}{4\pi} = 0.7892
\end{aligned}$$

The variance of the range of the original random variables is then

$$\mathrm{Var}[\mathbf{range}] = 0.7892\sigma^2$$

and the standard deviation of the range is 0.8884σ.

Recalling that we use $s_1=\text{range}/1.6926$ as an estimate of σ, the error in this estimate is

$$\mathbf{s}_1 - \sigma$$

and the expected squared value of this error is

$$\begin{aligned}
E[(\mathbf{s}_1-\sigma)^2] &= E[(\mathbf{range}/1.6926 - \sigma)^2] \\
&= \frac{1}{(1.6926)^2}E[(\mathbf{range} - E[\mathbf{range}])^2] \\
&= \frac{1}{(1.6926)^2}\mathrm{Var}[\mathbf{range}] \\
&= \frac{0.7892}{(1.6926)^2}\sigma^2 = 0.2755\sigma^2
\end{aligned}$$

How Good Are Other Estimates?

The more conventional method of computing the variance of a single observation from a sample of three numbers is to compute the mean of the three numbers

$$\bar{\mathbf{X}} = \frac{\sum_{i=1}^{3}\mathbf{X}_i}{3}$$

and the residuals from the mean

$$\mathbf{v}_i = \mathbf{X}_i - \bar{\mathbf{X}}$$

Then

$$\hat{\boldsymbol{\sigma}}^2 = \frac{\sum_{i=1}^{3}\mathbf{v}_i^2}{2}$$

is an unbiased estimate of σ^2.

It is shown in many texts on least-squares adjustments (e.g., Leick 1995, Sec. 4.9.2) that the sum of squares of residuals, divided by the true value of the variance of a single observation, is distributed as chi squared with $n-1$ degrees of freedom. Thus

$$\hat{\boldsymbol{\sigma}}^2 \sim \frac{\sigma^2}{2}\chi_2^2$$

From this

$$E[\hat{\boldsymbol{\sigma}}^2] = \frac{\sigma^2}{2}E[\chi_2^2] = \sigma^2, \quad \text{since } E(\chi_2^2)=2, \quad \text{and with } \mathrm{Var}(\chi_2^2)=4$$

$$\mathrm{Var}[\hat{\boldsymbol{\sigma}}^2] = \left(\frac{\sigma^2}{2}\right)^2\mathrm{Var}[\chi_2^2] = \left(\frac{\sigma^2}{2}\right)^2 2\times 2 = \sigma^4$$

Thus, the expected squared error of $\hat{\boldsymbol{\sigma}}^2$ as a estimate of σ^2 is σ^4.

```
BASELINE NAME:   galp  bkly
LLH  37 47 52.11771  278 50 52.03149     691.4034  NEW MON @ 2005.3222

 BASELINE NAME:   dobs  bkly
LLH  37 47 52.11762  278 50 52.03168     691.3831  NEW MON @ 2005.3222

 BASELINE NAME:   blkv  bkly
LLH  37 47 52.11736  278 50 52.03149     691.4094  NEW MON @ 2005.3222
```

Fig. 2. Portion of OPUS extended output

It is tempting to conclude from this that $\hat{\boldsymbol{\sigma}}$ is an unbiased estimate of σ and that its variance is σ^2. However, we find that this is not the case.

Let $\mathbf{y}=\Sigma_{i=1}^{3}\mathbf{v}_i^2/\sigma^2$. Then $\mathbf{y}$ is distributed as chi square with two degrees of freedom, and its probability density function is

$$f_{\mathbf{y}}(y)=\frac{1}{2\Gamma(1)}e^{-y/2},\quad \text{where the Gamma function } \Gamma(1)=0!=1$$

The estimated standard deviation is $\hat{\boldsymbol{\sigma}}=\sqrt{\sigma^2/2\mathbf{y}}$ and its expectation is

$$E[\hat{\boldsymbol{\sigma}}]=\frac{\sigma}{2\sqrt{2}}\int_{-\infty}^{\infty}y^{1/2}e^{-y/2}dy=\frac{\sigma}{2\sqrt{2}}\sqrt{2\pi}=\frac{\sqrt{\pi}}{2}=0.8662\sigma$$

This says that even though $\hat{\boldsymbol{\sigma}}^2$ is an unbiased estimate of σ^2, $\hat{\boldsymbol{\sigma}}$ underestimates σ. However, we can still find the expected squared error in this estimate as

$$E[(\hat{\boldsymbol{\sigma}}-\sigma)^2]=E[\hat{\boldsymbol{\sigma}}]-2E[\hat{\boldsymbol{\sigma}}\sigma]+E[\sigma^2]=E[\hat{\boldsymbol{\sigma}}^2]-2E[\hat{\boldsymbol{\sigma}}]\sigma+\sigma^2$$
$$=\sigma^2-2\times 0.8862\sigma\times\sigma+\sigma^2$$

So, finally,

$$E[(\hat{\boldsymbol{\sigma}}-\sigma)^2]=0.2275\sigma^2$$

This is slightly better than the expected squared error of $0.2755\sigma^2$ obtained when the range is used to estimate the standard deviation of a single observation. This seems intuitively correct, since it is based on all three observations rather than just two.

Example Using the OPUS Data Sheet

The OPUS solution report with extended output provides three different ways of estimating the uncertainty of the computed coordinates:

1. Using the range of the three solutions in latitude, longitude, and height. From Fig. 1, these are 0.011, 0.005, and 0.026 m. We find the standard deviation of the reported latitude, longitude, and height by dividing by 2.9317, yielding

$$\sigma_{\text{lat}}=0.004 \text{ m}$$

$$\sigma_{\text{long}}=0.002 \text{ m}$$

$$\sigma_{\text{height}}=0.009 \text{ m}$$

2. Using the residuals from the mean. Fig. 2 shows the three individual position determinations. The mean of these in each coordinate is the published International Terrestrial Reference Frame (ITRF) position, and the maximum minus minimum in each coordinate (converted to meters) is the reported range. From these, we can compute the residuals and estimate the standard deviation in each coordinate by

```
Covariance Matrix for the xyz OPUS Position (meters²).
    0.0000010156    -0.0000002797     0.0000001415
   -0.0000002797     0.0000143400    -0.0000008806
    0.0000001415    -0.0000008806     0.0000064200
```

Fig. 3. Another portion of OPUS extended output

$$\hat{\boldsymbol{\sigma}}=\sqrt{\frac{\sum_{i=1}^{3}\mathbf{v}_i^2}{2}}$$

This yields

$$\sigma_{\text{lat}}=0.005 \text{ m}$$

$$\sigma_{\text{long}}=0.003 \text{ m}$$

$$\sigma_{\text{height}}=0.014 \text{ m}$$

3. Using the covariance matrices of the individual baseline solutions. The single baseline processing engine used in OPUS produces a covariance matrix for each baseline. These are combined to produce the covariance matrix of the mean of the three individual coordinate determinations, rotated into east, north, and up coordinates, and reported in the OPUS extended output (Fig. 3). The square roots of the diagonal elements of this matrix also give the standard deviations of the reported coordinates. By this method we compute

$$\sigma_{\text{lat}}=0.001 \text{ m}$$

$$\sigma_{\text{long}}=0.003 \text{ m}$$

$$\sigma_{\text{height}}=0.003 \text{ m}$$

Concluding Remarks

The use of the range, or peak to peak errors, given on the data sheet may be advantageously used to estimate the standard deviations of the computed coordinates. This method of computing the standard deviations is almost as good as the more conventional method based on the sum of squares of residuals, and is also much more robust against the effects of a blunder in the data.

Notation

The following symbols are used in this paper:

Cov( ) = covariance of two random variables;
$E[\ ]$ = expected value;
Erf(x) = error function;
$F_{\mathbf{x}}(x)$ = cumulative probability function associated with the random variable $\mathbf{x}$, evaluated at the number x;
$f_{\mathbf{x}}(x)$ = probability density function associated with the random variable $\mathbf{x}$, evaluated at the number x;
G(x) = Gaussian probability function;
Var( ) = variance of a random variable; and
$\Gamma(x)$ = Gamma function.

Appendix. Joint Probability Function of the Minimum and Maximum of a Set of *N* Random Variables

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a set of n independent random variables and let $\mathbf{u} = \max\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and $\mathbf{v} = \min\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. The joint probability distribution function of $\mathbf{u}$ and $\mathbf{v}$ is found by determining the region of n-dimensional space such that $\{\mathbf{u} < u \& \mathbf{v} < v\}$ and then finding the probability density contained in this region (Papoulis 1965). Thus

$$\begin{aligned} F_{\mathbf{uv}}(u,v) &= \Pr\{\mathbf{u} < u \ \& \ \mathbf{v} < v\} \\ &= \Pr\{\mathbf{x}_1 < u \ \& \ \mathbf{x}_2 < u \ \& \ , \ldots, \ \& \ < \mathbf{x}_n < u\} \\ &\quad - \Pr\{v < \mathbf{x}_1 < u \ \& \ v < \mathbf{x}_2 < u \ \& \ , \ldots, \ \& \ v < \mathbf{x}_n < u\} \end{aligned}$$

The random variables $\mathbf{x}$ are all independent, so

$$F_{\mathbf{uv}}(u,v) = [\Pr\{\mathbf{x} < u\}]^n - [\Pr\{v < \mathbf{x} < u\}]^n$$

The probability in the second term is $F_{\mathbf{x}}(u) - F_{\mathbf{x}}(v)$ if $u > v$; otherwise it is zero. Thus

$$F_{\mathbf{uv}}(u,v) = \begin{cases} F_{\mathbf{x}}^n(u) - [F_{\mathbf{x}}(u) - F_x(v)]^n, & u > v \\ F_{\mathbf{x}}^n(u), & u < v \end{cases}$$

and the joint probability density function is

$$\begin{aligned} f_{\mathbf{uv}}(u,v) &= \frac{\partial^2 F_{\mathbf{uv}}(u,v)}{\partial u \partial v} \\ &= \begin{cases} n(n-1)[F_{\mathbf{x}}(u) - F_{\mathbf{x}}(v)]^{n-2} f_{\mathbf{x}}(u) f_{\mathbf{x}}(v), & u > v \\ 0, & u < v \end{cases} \end{aligned}$$

It is straightforward to verify the marginal distribution functions

$$F_{\mathbf{uv}}(u,\infty) = F_{\mathbf{x}}^n(u) = F_u(u)$$

$$F_{\mathbf{uv}}(\infty,v) = 1 - [1 - F_{\mathbf{x}}(v)]^n = F_{\mathbf{v}}(v)$$

and the marginal probability density functions

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\mathbf{uv}}(u,v)dv &= n(n-1)\int_{-\infty}^{u} [F_{\mathbf{x}}(u) - F_{\mathbf{x}}(v)]^{n-2} f_{\mathbf{x}}(u) f_{\mathbf{x}}(v) dv \\ &= nF_{\mathbf{x}}^{n-1}(u) f_{\mathbf{x}}(u) = f_{\mathbf{u}}(u) \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\mathbf{uv}}(u,v)du &= n(n-1)\int_{v}^{\infty} [F_{\mathbf{x}}(u) - F_{\mathbf{x}}(v)]^{n-2} f_{\mathbf{x}}(u) f_{\mathbf{x}}(v) du \\ &= n[1 - F_{\mathbf{x}}(v)]^{n-1} f_{\mathbf{x}}(v) = f_{\mathbf{v}}(v) \end{aligned}$$

References

Abramowitz, M., and Stegun, I., eds. (1977). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover, New York.

Leick, A. (1995). *GPS satellite surveying*, Wiley Interscience, New York.

Papoulis, A. (1965). *Probability, random variables, and stochastic processes*, McGraw-Hill, New York.

Weisstein, E. W. (undated). "Extreme value distribution," from *Mathworld—A Wolfram web resource*, ⟨http://mathworld.wolfram.com/ExtremeValueDistribution.htm⟩

Wilks, S. S. (1962). *Mathematical statistics*, Wiley, New York.

⟨http://www.ngs.noaa.gov/OPUS⟩

⟨http://www.ngs.noaa.gov/OPUS/Using_OPUS.html#accuracy⟩