On the distribution of the range of a sample

Steve Paik
Santa Monica College
(Dated: October 28, 2015)

Let $\{X_i\}$ be a sample of n independent and identically distributed random variables with common probability density function (pdf) f and cumulative distribution function (cdf) F. Let $U \equiv \max\{X_1, \ldots, X_n\}$ and $V \equiv \min\{X_1, \ldots, X_n\}$. Finally, let $Z \equiv U - V$.

I. THE DISTRIBUTION FOR THE SAMPLE MAXIMUM

It is easiest to think about the cdf:

$$\begin{aligned}
\Pr(U \leq u) &= \Pr[(X_1 \leq u) \cap \ldots \cap (X_n \leq u)] \\
&= \Pr(X_1 \leq u) \cdots \Pr(X_n \leq u) \\
&= \Pr(X \leq u)^n,
\end{aligned}$$

where we made use of both the independence and identically distributed properties. Thus, the cdf of U is the nth power of the cdf of X:

$$G(u) = F(u)^n.$$

The pdf is then

$$g(u) = nF(u)^{n-1} f(u).$$

II. THE DISTRIBUTION FOR THE SAMPLE MINIMUM

$$\begin{aligned}
\Pr(V > v) &= \Pr[(X_1 > v) \cap \ldots \cap (X_n > v)] \\
&= \Pr(X_1 > v) \cdots \Pr(X_n > v) \\
1 - \Pr(V \leq v) &= (1 - \Pr(X \leq v))^n.
\end{aligned}$$

So the cdf of V is

$$G(v) = 1 - (1 - F(v))^n.$$

The pdf is

$$g(v) = n(1 - F(v))^{n-1} f(v).$$

III. THE DISTRIBUTION FOR THE SAMPLE RANGE

$$\begin{aligned}
\Pr[(U < u) \cap (V > v)] &= \Pr[(v < X_1 < u) \cap \ldots \cap (v < X_n < u)] \\
&= \Pr(v < X_1 < u) \cdots \Pr(v < X_n < u) \\
&= \Pr(v < X < u)^n \\
&= \left(\int_v^u f(t)dt \right)^n \\
&= \left(\int_{-\infty}^u f(t)dt - \int_{-\infty}^v f(t)dt \right)^n \\
&= (F(u) - F(v))^n.
\end{aligned}$$

The joint pdf is obtained from this by taking a mixed second partial derivative $-\partial_u \partial_v$ (the minus sign is to account for the fact that we're differentiating minus the cdf for V),

$$g(u, v) = n(n-1)(F(u) - F(v))^{n-2} f(u) f(v).$$

The pdf for Z, which we shall call h, is obtained by integrating over the joint pdf with the constraint that $U - V = Z \geq 0$, which is a line in the (u, v) plane,

$$h(z) = \iint_R dudv\, g(u, v) \delta(u - v - z).$$

The integration region R is the intersection of the product of the two domains for u and v and the half-plane described by $v \leq u$.

Let us consider as our first example the uniform distribution on the interval $[0, L]$. This has pdf $f(x) = 1/L$ and cdf $F(x) = x/L$. The joint pdf is

$$h(z) = \int_0^L du \int_0^u dv\, n(n-1) \Big(\frac{u}{L} - \frac{v}{L} \Big)^{n-2} \Big(\frac{1}{L} \Big)^2 \delta(u - v - z).$$

Change variables to $(x, y) = (u - v, u + v)$. The Jacobian is $1/2$ and the triangular region in the (u, v) plane transforms to a larger half-diamond region in the first quadrant of the (x, y) plane bounded above by the line $y = 2L - x$ and bounded below by the line $y = x$. Therefore,

$$\begin{aligned}
h(z) &= \frac{n(n-1)}{2L^2} \int_0^L dx\, (x/L)^{n-2} \delta(x - z) \int_x^{2L-x} dy \\
&= \frac{1}{L} n(n-1)(z/L)^{n-2}(1 - z/L).
\end{aligned}$$

Since there is a peak the most probable value is given by

$$z_{\text{m.p.}}/L = \frac{n-2}{n-1}.$$

Note that $z_{\text{m.p.}} \to L$ as $n \to \infty$.

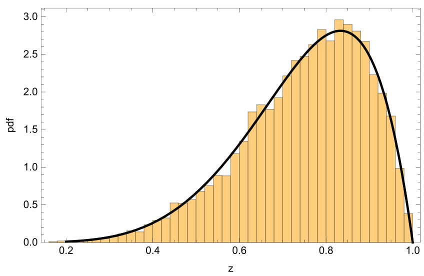

FIG. 1. PDF (black curve) for the sample range for sample sizes of $n = 7$ picked from the uniform distribution on the unit interval. The histogram shows the simulated results for 10,000 independent samples.

For the case of an unbounded random variable it certainly seems that the most probable value for z will keep increasing with n, for it is obvious that the larger a sample, the more extreme the endpoints can be. So it is not useful to take a large sample because the width of the distribution is generally a finite value. It is often the case that teachers ask students to estimate the standard deviation σ of a random variable by taking half the sample range. Therefore, the following question seems to be pertinent: Is there a value of n for which $z_{\text{m.p.}}$ is as close as possible to 2σ?

Consider a univariate distribution on $-\infty < \hat{x} < \infty$ described by the pdf $f(\hat{x})$ and the cdf $F(\hat{x})$. If we assume this distribution to be a location-scale family, then it is possible to prove that the same value of n is optimal for the entire family. Consider a location-scale transformation to some $X = a + b\hat{X}$ for $b > 0$. For instance, if $\hat{X}$ belongs to a standard normal distribution, then X would be distributed normally with mean a and standard deviation b. It follows that $F(x) = \Pr(X \le x) = \Pr(a + b\hat{X} \le x) = \Pr(\hat{X} \le (x-a)/b) = F((x-a)/b)$. Then, up to a factor of $n(n-1)$ which does not participate in the following transformations,

$$h(z) \propto \iint_{\mathbb{R}^2} \left[F\left(\frac{u-a}{b}\right) - F\left(\frac{v-a}{b}\right)\right]^{n-2} \delta(u-v-z) \times \frac{1}{b} f\left(\frac{u-a}{b}\right) \frac{1}{b} f\left(\frac{v-a}{b}\right) dudv.$$

Now set $\hat{u} = (u-a)/b$ and $\hat{v} = (v-a)/b$. Then

$$h(z)dz \propto \iint_{\mathbb{R}^2} [F(\hat{u}) - F(\hat{v})]^{n-2} \delta(b\hat{u} - b\hat{v} - z) f(\hat{u}) f(\hat{v}) d\hat{u} d\hat{v}.$$

But the delta function can be expressed as $b^{-1}\delta(\hat{u} - \hat{v} - \hat{z})$, where $\hat{z} = z/b$. Therefore,

$$h(z) = b^{-1} h(\hat{z}).$$

Now let us suppose that we have found the optimal value for n such that $\hat{z}_{\text{m.p.}}$, defined as the solution to $h'(\hat{z}_{\text{m.p.}}) = 0$, minimizes the quantity $|\hat{z}_{\text{m.p.}} - 2\hat{\sigma}|$ over all n. Here $\hat{\sigma}$ is the standard deviation associated to the $\hat{X}$. Then, for the new random variable X with associated standard deviation σ, we must minimize $|z_{\text{m.p.}} - 2\sigma| = b|\hat{z}_{\text{m.p.}} - 2\hat{\sigma}|$. But the same value of n accomplishes this since the scale factor b appears as an overall multiplicative constant. Looking back we see that this worked out the way it did because we chose to minimize a linear function $|z_{\text{m.p.}} - 2\sigma|$.

Next, let us motivate why the density $h(z)$ should have a peak at all. Since $0 < z < \infty$, consider the case of $z \ll 1$. Assuming that the cdf and pdf of the sample variable are smooth, then we can expand $F(v+z) \approx F(v) + zF'(v)$ and $f(v+z) \approx f(v) + zf'(v)$, to get

$$\begin{aligned} h(z) &= n(n-1) \int_{-\infty}^{\infty} dv\, (F(v+z) - F(v))^{n-2} f(v+z) f(v) \\ &\approx n(n-1) z^{n-2} \int_{-\infty}^{\infty} dv\, F'(v)^{n-2} (f(v) + zf'(v)) f(v) \\ &= n(n-1) z^{n-2} \int_{-\infty}^{\infty} dv\, f(v)^n + (n-1) z^{n-1} \big[f(v)\big]_{-\infty}^{\infty}. \end{aligned}$$

The second term vanishes and the first term's integral is finite. Thus, for small z, $h(z)$ behaves as z^{n-2}. This is an increasing function. For asymptotically large z, we may replace $F(v+z)$ by 1 since this does not result in the vanishing of the integral. However, we cannot replace $f(v+z)$ by 0. Let us suppose that the density is bounded as follows: for $x > x_1$,

$$f(x) < Ce^{-x},$$

for some constant C. Then asymptotically,

$$h(z) < Cn(n-1)e^{-z} \int_{-\infty}^{\infty} dv\, (1 - F(v))^{n-2} e^{-v} f(v),$$

so it vanishes at least as fast as an exponential decay. Therefore, at some finite z there must be one or more maxima of the function h. The locations of these maxima can only depend on n and the nature of the density f.

Let us apply these observations to the standard normal distribution with pdf $f(x) = e^{-x^2/2}/\sqrt{2\pi}$ and cdf $F(x) = \frac{1}{2}(1 + \text{erf}(x/\sqrt{2}))$. So

$$h(z) = \frac{n(n-1)}{\pi 2^{n-1}} e^{-z^2/2} \times \int_{-\infty}^{\infty} dv \left[\text{erf}\left(\frac{v+z}{\sqrt{2}}\right) - \text{erf}\left(\frac{v}{\sqrt{2}}\right)\right]^{n-2} e^{-v^2 - vz}.$$

This integral needs to be evaluated numerically for a given n and a list of points z. Doing so shows that $h(z)$ has a single hump for any n. See Fig. 2. Recall that $\hat{\sigma} = 1$. We note that $|z_{\text{m.p.}} - 2|$ is smallest for $n = 5$, although $n = 4$ is not far away either. See Table I.

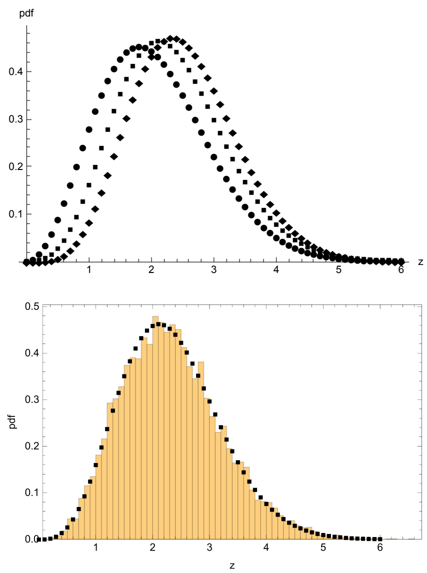

FIG. 2. Top: PDF's for the sample range for sample sizes of $n = 4$ (circles), 5 (squares), and 6 (diamonds) picked from the standard normal distribution. Bottom: The pdf for $n = 5$ and a histogram showing the simulated results for 10,000 independent samples.

n	$z_{\mathrm{m.p.}}$
4	1.806
5	2.113
6	2.343

TABLE I. The location of the maxima of the sample range density function. These are the most probable values for the sample range.

This suggests that if one suspects that the iid variables are normally distributed with any mean and any variance, then an optimal strategy to estimate (twice) the standard deviation is to use the range of just five measurements.

Suppose one takes a sample of size $n = 5$ from iid normal variables. What is the probability that the sample range is within $p = 10\%$ of (twice) the actual standard deviation? For a location-scale family distribution, we need to compute

$$\Pr\Big(\Big|\frac{z-2\sigma}{2\sigma}\Big| < p\Big) = \Pr\Big(\Big|\frac{b\hat{z}}{2b\hat{\sigma}} - 1\Big| < p\Big).$$

We see that by phrasing the question in terms of a percent deviation, the scale factor b drops out. In particular, for the normal distribution we must calculate

$$\begin{aligned}\Pr(|\hat{z}/2-1| < p) &= \Pr(2(1-p) < \hat{z} < 2(1+p)) \\ &= \int_{2-2p}^{2+2p} h(z)dz \\ &= 0.18\end{aligned}$$

So there is an 18% chance that the sample range is within 10% of the actual value. By taking $p = 0.3$, it turns out that there is a 51% chance that the sample range is within 30% of the actual value. By taking $p = 0.75$, it turns out that there is a 90% chance that the sample range is within 75% of the actual value.

IV. ON THE EXPONENTIAL DISTRIBUTION

The exponential distribution is a continuous probability distribution that describes the time between events in a Poisson process. We denote these time intervals by X. The renewal assumption of the Poisson process fixes the distribution of X up to a single positive parameter. This is because the memoryless property is equivalent to the law of exponents.

The density function is $f(t) = re^{-rt}$. Some moments are $\mathbb{E}(X) = 1/r$ and $\mathrm{var}(X) = 1/r^2$. This shows that, on average, there are $1/r$ time units between events.

A. Distribution of the sample range

We must evaluate

$$h(z) = \int_0^\infty du \int_0^u dv\, g(u,v)\delta(u-v-z)$$

where

$$g(u,v) = r^2 n(n-1)(e^{-rv} - e^{-ru})^{n-2} e^{-ru-rv}.$$

The delta function enforces the constraint that $v = u - z$. However, for $u < z$, this implies $v < 0$ which is not part of the allowed integration region. Thus, the delta function vanishes identically for $0 \le u \le z$. Thus,

$$\begin{aligned}h(z) &= r^2 n(n-1) \int_z^\infty du\, (e^{-ru+rz} - e^{-ru})^{n-2} e^{-2ru+rz} \\ &= r^2 n(n-1) e^{rz}(e^{rz}-1)^{n-2} \int_z^\infty du\, e^{-nru} \\ &= r^2 n(n-1) e^{rz}(e^{rz}-1)^{n-2} \frac{e^{-nrz}}{nr} \\ &= (n-1) r e^{-rz}(1-e^{-rz})^{n-2}.\end{aligned}$$

The cdf is simply

$$H(z) = (1-e^{-rz})^{n-1}.$$

The peak of the pdf occurs when $h' = 0$. So $z_{\mathrm{m.p.}} = r^{-1}\ln(n-1)$. Note that, in terms of the cdf, this is where it has an inflection point ($H'' = 0$). Minimizing $|z_{\mathrm{m.p.}} - \sqrt{\mathrm{var}(X)}| = r^{-1}|\ln(n-1) - 1|$ over n, we find that n is the nearest integer to $1+e$. So $n = 4$ is optimal. Note that we could have reached this conclusion using $r = 1$ from the outset since the exponential distribution is an example of a scale family. However, the computations were simple enough to keep the r arbitrary.

B. Distribution of the sample mean

Suppose X_1, X_2, X_3, and X_4 are independent random variables exponentially distributed with $r = 1$. What then is the distribution for the sum $Y \equiv X_1 + X_2 + X_3 + X_4$? We can either find the cdf:

$$F(y) = \int_{[0,\infty)^4} dx_1 dx_2 dx_3 dx_4 \, e^{-x_1-x_2-x_3-x_4} \times \Theta(y - x_1 - x_2 - x_3 - x_4),$$

where $\Theta(x)$ is the unit step function and is 0 for $x < 0$ and 1 for $x \geq 0$; or the pdf:

$$f(y) = \int_{[0,\infty)^4} dx_1 dx_2 dx_3 dx_4 \, e^{-x_1-x_2-x_3-x_4} \times \delta(y - x_1 - x_2 - x_3 - x_4).$$

We will compute the pdf directly and, if we want, we can always integrate to obtain the cdf. The pdf is simpler to find because, by doing the integral over, say x_4, we are left with a volume integral of a triangular pyramid,

$$\begin{aligned} f(y) &= e^{-y} \int_{\substack{x_1+x_2+x_3<y \\ x_i \geq 0}} dx_1 dx_2 dx_3 \\ &= \frac{y^3}{6} e^{-y}, \quad y \geq 0. \end{aligned}$$

For arbitrary $r > 0$, we can rescale $y \to ry$ so that $f(y)dy \to \frac{1}{6} r^4 y^3 e^{-ry} dy$. The pdf of the sample mean $W \equiv Y/4$ is obtained by letting $y = 4w$. Then $f(w)dw = f(y)dy$ so $f(w) = \frac{128}{3} r^4 w^3 e^{-4rw}$. This has a peak at $w_{\text{m.p.}} = (3/4) r^{-1}$.

For arbitrary n and $r = 1$, it is obvious from scaling that $f(y) = y^{n-1} e^{-y} / \Gamma(n)$. Therefore, $f(w) = \frac{n^n}{\Gamma(n)} w^{n-1} e^{-nw}$. The peak occurs at $w_{\text{m.p.}} = (n-1)/n$. In the limit $n \to \infty$, the density approaches a Dirac delta function centered at $w = 1$.

For $n = 4$, the most probable value of the sample range is only about 10% off from the standard deviation of the underlying distribution, but the most probable value of the sample mean is 25% off. However, the density of sample means is narrower and taller than that of the sample ranges. See Fig. 3. What is the probability that the sample range or sample mean will be within $p = 10\%$ of the actual value? Numerical integration shows that they are 9% and 16%, respectively.

C. Joint distribution of the sample maximum, minimum, and total

For the exponential distribution, $\mathbb{E}(X) = \sqrt{\text{var}(X)} = \text{sd}(X) = 1/r$. Therefore, it one wants to estimate the standard deviation, then one should take as large a sample size as possible. The law of large numbers

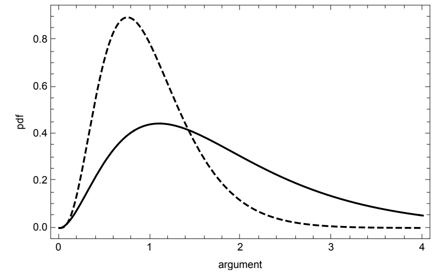

FIG. 3. Pdfs for the sample range (solid) and sample mean (dashed) for 4 identically exponentially distributed random variables with $r = 1$.

states that the sample mean converges to the distribution mean as the sample size increases. This would be the most straightforward way to estimate the rate parameter. However, there is a nice economy to keeping n as small as possible. And student bias often makes it difficult to take more measurements than the number of people in a group, which is typically four. As we saw in the last section, with $n = 4$, there is a better chance that the sample mean matches the distribution mean rather than the sample range matching the mean. An interesting question is whether a simple rule-of-thumb can be found for the other degrees of freedom in the data set that strengthens the likelihood that the sample mean estimates the true mean. Or, in keeping with the theme of this note which is to use the sample range as a substitute for the standard deviation, can we strengthen the likelihood that the sample range in some way bounds the standard deviation above or below? We will find an answer to this latter question.

It is of interest to compute the joint pdf for the sample maximum $U \equiv \max\{X_i\}$, sample minimum $V \equiv \min\{X_i\}$, and sample total $Y \equiv \sum_i X_i$. Assume each X_i is exponentially distributed with $r = 1$. Consider the cumulative probability

$$\begin{aligned} J(U, V, Y) \equiv \Pr[&(V \leq X_1 \leq U) \cap \ldots \cap (V \leq X_4 \leq U) \\ &\cap (X_1 + X_2 + X_3 + X_4 < Y)]. \end{aligned}$$

It is defined by the integral

$$J(u, v, y) = \int_{[v,u]^4} d^4x \, e^{-\sum_i x_i} \Theta(y - \sum_i x_i).$$

Since one of the four variables must be v and another one must be u, it follows that

$$J > 0 \quad \text{only for} \quad u + 3v < y < 3u + v.$$

We will take the mixed partial derivatives $\partial_u(-\partial_v)\partial_y$ in order to obtain the joint density j. As before, let us take the partial derivative with respect to y and work with

$$\partial_y J = \int_{[v,u]^4} d^4x \, e^{-\sum_i x_i} \delta(y - \sum_i x_i).$$

Doing the integral over, say, x_4, results in

$$\partial_y J = e^{-y} \int_{\substack{y-u<x_1+x_2+x_3<y-v \\ v\le x_i\le u}} dx_1 dx_2 dx_3.$$

Note the complicated bounds on the three remaining variables. The reason they exist is because otherwise no choice for x_4 could satisfy the constraint that $\sum_i x_i = y$. This volume integral is tedious to work out. We do so in the appendix and quote the result of taking the remaining mixed partial derivatives here:

$$-\partial_u \partial_v \partial_y J = e^{-y} \begin{cases} 12(-u-3v+y), & y < 2u+2v \\ 12(3u+v-y), & y > 2u+2v \end{cases}.$$

We note that both expressions agree at $y = 2u+2v$. The complete expression for the joint density is

$$j(u,v,y) = 12e^{-y} \begin{cases} (-u-3v+y), & u+3v < y < 2u+2v \\ (3u+v-y), & 2u+2v < y < 3u+v \\ 0, & \text{else} \end{cases}.$$

We should be able to check that this has probability unity when fully integrated and that the marginal distributions are the same as what we previously derived. Indeed, using *Mathematica* we checked that

$$\int_0^\infty du \int_0^u dv \int_{u+3v}^{3u+v} dy\, j(u,v,y) = 1.$$

Here it is not necessary to specify the limits on y so precisely. We could have set the limits as $y \in [0,\infty)$. Note however, that $u > v$ so we set the upper limit on v as u. We also checked using *Mathematica* that for $u > z > 0$,

$$\int_z^\infty du \int_0^u dv\, \delta(v-(u-z)) \int_{u+3v}^{3u+v} dy\, j(u,v,y) = 3e^{-z}(1-e^{-z})^2.$$

This is the density of the sample range. To get the marginal density for the sample mean we must allow the y limits to span $[0,\infty)$ and pull this integral to the outside. Then, for a given y, the integrand is supported only over the regions of the (u,v) plane defined implicitly by

$$R_1(y) \equiv v < u \ \cap \ u+3v < y \ \cap \ 2u+2v > y$$

over which the weight $12e^{-y}(-u-3v+y)$ is to be integrated, and

$$R_2(y) \equiv v < u \ \cap \ 2u+2v < y \ \cap \ 3u+v > y$$

over which the weight $12e^{-y}(3u+v-y)$ is to be integrated. These regions have finite area so the integral converges. Using *Mathematica*, we find that the first region gives $\frac{1}{8}y^3e^{-y}$ and the second region gives $\frac{1}{24}y^3e^{-y}$. Hence,

$$\int_0^\infty dy \int_{R_1(y)\cup R_2(y)} dudv\, j(u,v,y) = \tfrac{1}{6}y^3e^{-y}.$$

Let calculate the covariance between the jointly distributed real-valued random variables $Z \equiv U - V$ and Y. Using *Mathematica*,

$$\begin{aligned} \mathrm{cov}(Y,Z) &= \mathbb{E}[(Y-\mathbb{E}(Y))(Z-\mathbb{E}(Z))] \\ &= \mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z) \\ &\approx 1.8333. \end{aligned}$$

The fact that this is greater than zero shows that Y and Z are positively correlated. In other words, Y and Z are not independent — the joint pdf is not a product of their marginal pdf's. The correlation is

$$\mathrm{cor}(Y,Z) = \frac{\mathrm{cov}(Y,Z)}{\mathrm{sd}(Y)\mathrm{sd}(Z)} \approx 0.7857.$$

The fact that they are correlated is pretty obvious considering that the variables X_i, Z and Y are positive.

For $n = 4$ (and still for $r = 1$) the probability that $\mathrm{sd}(X) = 1 < Z$ is given by

$$\Pr(Z > 1) = \int_1^\infty dz\, 3e^{-z}(1-e^{-z})^2 \approx 0.7474.$$

There is about a 75% chance that the standard deviation lies below the sample range as is obvious from staring at the solid curve of Fig. 3. This serves as a useful *probable upper bound* on the standard deviation.

In fact, it is best to use $Y/4$ as the estimate for the standard deviation.

We ask whether a comparison of the relative sizes of $Y/4$ and Z can enhance or degrade the likelihood that the standard deviation lies below the sample range?

$$\begin{aligned} \Pr(Z > 1 | Z > Y/4) &= \frac{\Pr(Z > 1 \cap Z > Y/4)}{\Pr(Z > Y/4)} \\ &\approx \frac{0.7251}{0.9063} \\ &\approx 0.80. \end{aligned}$$

So simply by checking and finding that the sample mean is less than the sample range we are now about 80% confident that the standard deviation is less than the sample range. This represents a 5% enhancement in our probable upper bound for the standard deviation.

On the other hand,

$$\begin{aligned} \Pr(Z > 1 | Z < Y/4) &= \frac{\Pr(1 < Z < Y/4)}{\Pr(Z < Y/4)} \\ &\approx \frac{0.0223}{0.0938} \\ &\approx 0.24. \end{aligned}$$

So if it is found that the sample mean is greater than the sample range, then we are only about 24% confident that the standard deviation is less than the sample range. So it's more likely that the standard deviation is above the sample range. That makes sense since the sample mean tends to track the standard deviation. This situation won't happen that often — only about once out of every ten samples — but when it does happen it suggests that the sample range is a *probable lower bound* for the standard deviation.

Lastly, we remark that the above discussion for the rate $r = 1$ is valid for arbitrary $r > 0$. It is easy to go back and see that the joint pdf would be

$$j_r = r^3 j(ru, rv, ry).$$

Unit probability is preserved because one can rescale integration variables as $\tilde{u} \equiv ru$, $\tilde{v} \equiv rv$, and $\tilde{y} \equiv ry$. Then the integral is identical to the one considered above. Also, any conditional probability in which a constraint like $Z > 1/r$ appears is identical to the $r = 1$ case. Observe that $Z > 1/r \implies rZ > 1 \implies \tilde{Z} > 1$. So constraints given by an inequality which is homogeneous in the random variables may be studied using the $r = 1$ case.

For the sake of reference, a probability like $\Pr(Z > 1 \cap Z > Y/4)$ is evaluated in *Mathematica* as

```
NIntegrate[j[u,v,y]*Boole[u-v>1&&u-v>y/4],{u,0,40.0},{v,0,u},{y,u+3v,3u+v}]
```

Appendix A: Evaluation of implicitly defined volume

Let

$$I \equiv \int_{\substack{y-u<x_1+x_2+x_3<y-v \\ v\le x_i \le u}} dx_1 dx_2 dx_3.$$

This may be written

$$I = \int_{[v,u]^2} dx_1 dx_2 \int_{\max(v,y-u-x_1-x_2)}^{\min(u,y-v-x_1-x_2)} dx_3.$$

It turns out the limits are decided by the value of x_1+x_2. If $x_1+x_2 < y-u-v$, then the lower limit is $y-u-x_1-x_2$ and the upper limit is u. Of course, one must require that $y-u-x_1-x_2 < u$ which implies the further restriction that $x_1+x_2 > y-2u$. However, if $x_1+x_2 > y-u-v$, then the lower limit is v and the upper limit is $y-v-x_1-x_2$. Now one must require that $v < y-v-x_1-x_2$ which implies that $x_1+x_2 < y-2v$. So

$$I = \int_{\substack{y-2u<x_1+x_2<y-u-v \\ v<x_i<u}} dx_1 dx_2 [u-(y-u-x_1-x_2)] + \int_{\substack{y-u-v<x_1+x_2<y-2v \\ v<x_i<u}} dx_1 dx_2 [y-v-x_1-x_2-v].$$

Call the integrals A and B, respectively. The integration region is some portion of a rectangle $[v,u]^2$ in the (x_1,x_2) plane. Since the constraints define a 45-degree line, it matters what the value of y is.

Case *(i)*: $y < 2u+2v$

Therefore, $y-u-v < u+v$ and $y-2u < 2v$. So for the A integral, the integration region is a triangle that starts flush with the bottom left corner of the rectangle and stops below the diagonal:

$$A = \int_v^{y-u-2v} dx_1 \int_v^{-x_1+y-u-v} dx_2 [-y+2u+x_1+x_2].$$

The defining constraint of this case also implies that $y-2v < 2u$. Moreover, $y > u+3v$ implies that $y-2v > u+v$. So for the B integral, the integration region is a diagonal strip that starts below the diagonal and ends above the diagonal:

$$B = \int_v^{y-u-2v} dx_1 \int_{-x_1+y-u-v}^{u} dx_2 [y-2v-x_1-x_2] + \int_{y-u-2v}^{u} dx_1 \int_v^{-x_1+y-2v} dx_2 [y-2v-x_1-x_2].$$

These integrals are straightforward. We find

$$I = \tfrac{2}{3}u^3 + 6u^2v + 18uv^2 + \tfrac{22}{3}v^3 + (-2u^2-12uv-10v^2)y + (2u+4v)y^2 - \tfrac{1}{2}y^3.$$

Case *(ii)*: $y > 2u+2v$

This implies that $y-2u > 2v$ and $y-u-v > u+v$. Furthermore, $y < 3u+v$ implies $y-2u < u+v$ and $y-u-v < 2u$. So for the A integral, the integration region is a diagonal strip that starts below the diagonal and ends above the diagonal:

$$A = \int_v^{y-2u-v} dx_1 \int_{-x_1+y-2u}^{u} dx_2 [-y+2u+x_1+x_2] + \int_{y-2u-v}^{u} dx_1 \int_v^{-x_1+y-u-v} dx_2 [-y+2u+x_1+x_2].$$

The defining constraint of this case also implies that $y-2v > 2u$. Thus, the integration region for B is the upper right triangle:

$$B = \int_{y-2u-v}^{u} dx_1 \int_{-x_1+y-u-v}^{u} dx_2 [y-2v-x_1-x_2].$$

These integrals straightforwardly evaluate to

$$I = -\tfrac{22}{3}u^3 - 18u^2v - 6uv^2 - \tfrac{2}{3}v^3 + (10u^2+12uv+2v^2)y + (-4u-2v)y^2 + \tfrac{1}{2}y^3.$$