1 Boas, p. 755, problem 15.6-13

Given a joint distribution function $f(x,y)$, show that $E(x+y) = E(x)+E(y)$ and $\mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y) + 2\,\mathrm{Cov}(x,y)$.

We only need to remember the definitions:

$$E(x) = \bar{x} = \int_{-\infty}^{\infty} dx\, dy\, x f(x,y)\,, \qquad E(y) = \bar{y} = \int_{-\infty}^{\infty} dx\, dy\, y f(x,y)\,, \tag{1}$$

$$E(x+y) = \int_{-\infty}^{\infty} dx\, dy\, (x+y) f(x,y) = \int_{-\infty}^{\infty} dx\, dy\, x f(x,y) + \int_{-\infty}^{\infty} dx\, dy\, y f(x,y) = E(x)+E(y)$$

$$\mathrm{Var}(x) = \int_{-\infty}^{\infty} dx\, dy\, (x-\bar{x})^2 f(x,y)\,, \mathrm{Var}(y) = \int_{-\infty}^{\infty} dx\, dy\, (y-\bar{y})^2 f(x,y)\,, \tag{2}$$

$$\mathrm{Cov}(x,y) = \int_{-\infty}^{\infty} dx\, dy\, (x-\bar{x})(y-\bar{y}) f(x,y)\,, \tag{3}$$

$$\begin{aligned}
\mathrm{Var}(x+y) &= \int_{-\infty}^{\infty} dx\, dy\, [(x+y) - E(x+y)]^2 f(x,y) = \int_{-\infty}^{\infty} dx\, dy\, [(x+y) - (\bar{x}-\bar{y})]^2 f(x,y) \\
&= \int_{-\infty}^{\infty} dx\, dy\, [(x-\bar{x}) + (y-\bar{y})]^2 f(x,y) \\
&= \int_{-\infty}^{\infty} dx\, dy\, [(x-\bar{x})^2 + (y-\bar{y})^2 + 2(x-\bar{x})(y-\bar{y})] f(x,y) = \mathrm{Var}(x) + \mathrm{Var}(y) + 2\,\mathrm{Cov}(x,y)
\end{aligned}$$

2 Boas, p. 756, problem 15.6-16

If x and y are independent, then $\mathrm{Cov}(x,y) = 0$. The converse is not always true, that is, if $\mathrm{Cov}(x,y) = 0$ it is not necessarily true that the joint distribution function is of the form $f(x,y) = g(x)h(y)$. For example, suppose $f(x,y) = (3y^2 + \cos x)/4$ on the rectangle $-\frac{\pi}{2} < x < \frac{\pi}{2}$, $-1 < y < 1$, and $f = 0$ elsewhere. Show that $\mathrm{Cov}(x,y) = 0$ but x and y are not independent. Can you construct some more examples?

We have

$$E(x) = \int_{-\pi/2}^{\pi/2} x dx \int_{-1}^{1} dy\, \frac{3y^2 + \cos x}{4} = \int_{-\pi/2}^{\pi/2} \frac{2+\cos x}{4}\, x dx = 0 \tag{4}$$

$$E(y) = \int_{-1}^{1} y dy \int_{-\pi/2}^{\pi/2} dx\, \frac{3y^2 + \cos x}{4} = \int_{-1}^{1} \frac{3\pi y^2 + 2}{4}\, y dy = 0 \tag{5}$$

$$\mathrm{Cov}(x,y) = \int_{-\pi/2}^{\pi/2} dx \int_{-1}^{1} dy\, (x-\bar{x})(y-\bar{y}) f(x,y) = \int_{-\pi/2}^{\pi/2} dx \int_{-1}^{1} dy\, xy \frac{3y^2 + \cos x}{4} = 0 \tag{6}$$

In each of the final steps, we have used the fact that $f(x,y) = f(-x,y) = f(x,-y) = f(-x,-y)$, and noted that the final integration is of an odd function taken over an integral symmetric about the origin. Hence, the integrals above all vanish.

We can give more examples of this type: for simplicity, let us use distribution functions which are even over a domain symmetric around the origin, so that $\bar{x} = \bar{y} = 0$. Now, for any even function $g(x)$, $h(y)$ we can define

$$f(x,y) = g(x) + h(y)\,, \quad f,\ g \text{ even}, \quad -a < x < a\,,\ -b < y < b \tag{7}$$

such that

$$\bar{x} = 0,\ \bar{y} = 0\,, \quad \mathrm{Cov}(x,y) = \int_{-a}^{a} dx \int_{-b}^{b} dy\, xy(g(x) + h(y)) = 0 \tag{8}$$

Note that $f(x,y)$ cannot be expressed as a product of a function of x times a function of y; hence x and y are not independent. An example is

$$f(x,y) \propto 5\cosh x + x^2 + y^4\,, \tag{9}$$

where the constant of proportionality is adjusted such that the $\int dx \int dy f(x,y) = 1$.

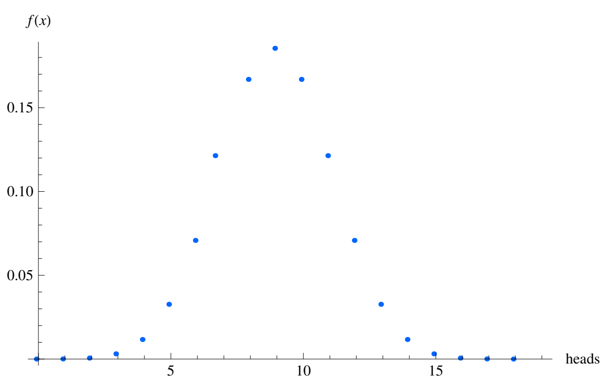
3 Boas, p. 760, problem 15.7-4

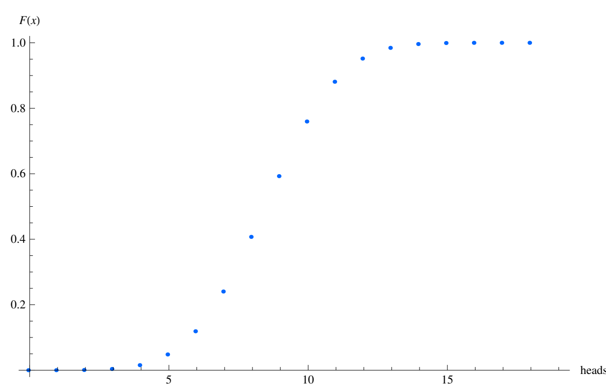
(a) Write the probability density function $f(x)$ for the probability of of x heads in $n = 18$ tosses of a coin and plot a graph of $f(x)$. Also plot the corresponding cumulative distribution function $F(x)$.

The probability density function for x heads is

$$f(x) = C(18,x)\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{18-x} = C(18,x)2^{-18} = \frac{18!}{x!(18-x!)}2^{-18} \tag{10}$$

The plot for the probability density and for the cumulative density are





(b) Plot a graph of $nf(x)$ as a function of $\frac{x}{n}$

It takes the same form of the graph of $f(x)$ above, except on the interval $[0,1]$. This form of the graph is more useful when comparing different n values. In this problem, we are only considering $n = 18$, so there is not much more to say.

(c) What is the probability of 7 heads? Of at most 7 heads? Of at least 7 heads? What is the most probable number of heads? The expected number of heads?

The probability of 7 heads is $\frac{18!}{7!11!}2^{-18} = 0.1214$. The probability of at most 7 heads is given by the cumulative function calculated in 7, $F(7) = 0.24$. The probability of having at least 7 heads is one minus the probability of having at most 7 heads, plus the probability of having 7 heads, that is, $0.76+0.12 = 0.88$. The most probable number of heads is 9 heads, the peak of the distribution. The expected number of heads is

$$\sum_{k=0}^{18} kC(18,k)2^{-18} = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) kp^k q^{n-k} = p\frac{\partial}{\partial p}\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} = p\frac{\partial}{\partial p}(p+q)^n = pn(p+q)^{n-1} = np = 9$$

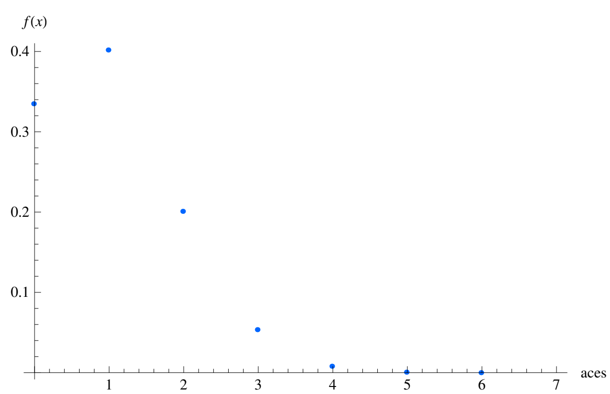
4 Boas, p. 760, problem 15.7-5

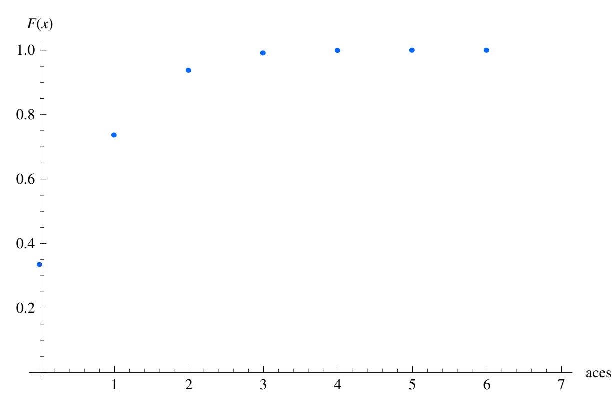
Write the formula for the binomial density function for the case $n = 6$, $p = \frac{1}{6}$ and plot it. Also plot the cumulative distribution function $F(x)$. What is the probability of at least 2 aces out of 6 tosses of a die?

The probability density function for x aces is

$$f(x) = C(6,x)\left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{6-x} = \frac{6!}{x!(6-x!)}5^{6-x}6^{-6} \tag{11}$$

The plot for the probability density and for the cumulative density are





The probability for 1 ace can be read in the graph: it has a probability of 0.4019; the probability for at least 2 aces is given by

$$1 - \text{probability of getting 0 or 1 ace} = 1 - \left(\frac{5}{6}\right)^6 - 6\frac{1}{6}\left(\frac{5}{6}\right)^5 = 0.2632. \tag{12}$$

5 Boas, p. 765, problem 15.8-3

The probability density for the x component of a the velocity of a molecule of an ideal gas is proportional to $e^{-mv^2/2kT}$. By comparing this to the standard expression of the normal distribution, eq. 8.1 of Boas, p761, find the mean and standard deviation of v and write the probability density function $f(v)$.

The probability density is written as

$$f(v) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(v-\mu)^2/2\sigma^2} \tag{13}$$

If we compare this to $e^{-mv^2/2kT}$, we see that

$$\mu = 0\,, \qquad \sigma^2 = \frac{kT}{m}\,, \qquad f(v) = \sqrt{\frac{m}{2\pi kT}}\, e^{-mv^2/(2kT)} \tag{14}$$

6 Boas, p. 765, problem 15.8-8

Start with the binomial distribution function: we want an approximation of it for large n.

We start with

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \tag{15}$$

using Stirling's formula, $n! \approx n^n e^{-n}\sqrt{2\pi n}$, we have

$$\begin{aligned} f(x) &\approx \frac{n^n e^{-n}\sqrt{2\pi n}}{x^x e^{-x}\sqrt{2\pi x}(n-x)^{n-x} e^{-(n-x)}\sqrt{2\pi(n-x)}} p^x q^{n-x} \\ &= (p/x)^x (q/(n-x))^{n-x} n^n \sqrt{\frac{n}{2\pi x(n-x)}} \\ &= \left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} \sqrt{\frac{n}{2\pi x(n-x)}} \end{aligned} \tag{16}$$

If we now define $\delta = x - np$, we have $x = \delta + np$, $n - x = nq - \delta$, we can express the first two terms in another way:

$$\ln\frac{np}{x} = \ln\frac{np}{np+\delta} = -\ln\left(1+\frac{\delta}{np}\right)\,, \qquad \ln\frac{nq}{n-x} = \ln\frac{nq}{nq-\delta} = -\ln\left(1-\frac{\delta}{nq}\right) \tag{17}$$

Then, using the expansion $\ln(1+x) = x - \frac{1}{2}x^2 + \mathcal{O}(x^3)$, we have

$$\begin{aligned} \ln\left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} &= x\ln\frac{np}{x} + (n-x)\ln\frac{nq}{n-x} \\ &= -(\delta+np)\left[\frac{\delta}{np} - \frac{1}{2}\frac{\delta^2}{n^2p^2} + \mathcal{O}\left(\frac{\delta^3}{n^3}\right)\right] - (nq-\delta)\left[-\frac{\delta}{nq} - \frac{1}{2}\frac{\delta^2}{n^2q^2} + \mathcal{O}\left(\frac{\delta^3}{n^3}\right)\right] \\ &= -\delta\left[1 + \frac{1}{2}\frac{\delta}{np} - 1 + \frac{1}{2}\frac{\delta}{nq} + \mathcal{O}\left(\frac{\delta^2}{n^2}\right)\right] = -\frac{\delta^2}{2npq} + \mathcal{O}\left(\frac{\delta^3}{n^2}\right)\,, \end{aligned} \tag{18}$$

so that the first two terms in (16) can be written as

$$\left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} = e^{-\delta^2/2npq}\left[1 + \mathcal{O}\left(\frac{\delta^3}{n^2}\right)\right]\,, \tag{19}$$

while the square root can be approximated with

$$\sqrt{\frac{n}{2\pi x(n-x)}} = \sqrt{\frac{n}{2\pi(np+\delta)(nq-\delta)}} = \sqrt{\frac{1}{2\pi npq}}\left[1 + \mathcal{O}\left(\frac{\delta}{n}\right)\right]\,. \tag{20}$$

In class, I argued that x should differ from the mean $\mu = np$ by a number of standard deviations, $\sigma = \sqrt{npq}$, that does not grow with n. Since $x = np + \delta$, this means that at worst, $\delta \sim \mathcal{O}(\sqrt{n})$. In this case, both $\mathcal{O}(\delta^3/n^2)$ and $\mathcal{O}(\delta/n)$ in (19) and (20) behave as $\mathcal{O}(1/\sqrt{n})$ as $n \to \infty$. Hence, the binomial probability density function can been written as

$$f(x) = \frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq} \left[1 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right] , \tag{21}$$

which is the normal distribution with parameters $\mu = np$, $\sigma^2 = npq$, up to corrections that vanish as $n \to \infty$. Indeed μ and σ of the normal approximation are identical to mean value and the standard deviation of the original binomial distribution. The normal approximation holds for x within some number of standard deviations of the average value np, where this number is of $\mathcal{O}(1)$ as $n \to \infty$, which corresponds to the central part of the bell curve. Since f is small anyway in other parts of the domain, we can ignore the fact that our approximation may not be good there.

7 Boas, p. 766, problem 15.8-15

Use the binomial distribution and the corresponding normal approximation to find the probabilities of having exactly 195 tails in 400 tosses of a coin.

Using the binomial distribution, with $n = 400$, $x = 195$, we have

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{n!}{x!(n-x)!} 2^{-n} = \frac{400!}{195!205!} 2^{-400} = 0.035195 \tag{22}$$

Using the normal approximation, we have $\mu = np = 200$, $\sigma = \sqrt{npq} = 10$,

$$f(x) = \frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq} = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} = \frac{1}{10\sqrt{2\pi}} e^{-25/200} = 0.035206 \tag{23}$$

The difference between the two estimates is only 1.1×10^{-5}.

8 Boas, p. 766, problem 15.8-25

An instructor who grades "on the curve" computes the mean and the standard deviation of the grades, and then, assuming a normal distribution with this μ and σ, sets the border lines between the grades at: C from $\mu - \frac{1}{2}\sigma$ to $\mu + \frac{1}{2}\sigma$, B from $\mu + \frac{1}{2}\sigma$ to $\mu + \frac{3}{2}\sigma$, A from $\mu + \frac{3}{2}\sigma$ up, etc. Find the percentages of the students receiving the various grades. Where should the border lines be set to give the percentages: A and F, 10%; B and D, 20%; C, 40%.

To find the percentages we must integrate the normal distribution in the various intervals: we will use the fact that the distribution is symmetric and so are the cuts used by the instructor, so that the percentages relative to A and F will be the same, and so the ones relative to B and D. First let us compute the percentage of students with a C:

$$P_C = \int_{\mu-\frac{1}{2}\sigma}^{\mu+\frac{1}{2}\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx = \int_{-\frac{1}{2}\sigma}^{\frac{1}{2}\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} = \tag{24}$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\frac{1}{2\sqrt{2}}} e^{-y^2} dy = \mathrm{erf}\left(\frac{1}{2\sqrt{2}}\right) = 0.3829 = 38.29\% \tag{25}$$

In the first line first we have shifted the integral, then we have changed variable $y = x/\sigma$ and finally we have used the fact that the integrand is even to recognize the error function, $\mathrm{erf}(x)$. The values of the error function are tabulated.

For the other grades the manipulations are the same:

$$P_B + P_D = \frac{2}{\sqrt{\pi}} \int_{\frac{1}{\sqrt{2}}}^{\frac{3}{2\sqrt{2}}} e^{-y^2} dy = \mathrm{erf}(\tfrac{3}{2\sqrt{2}}) - \mathrm{erf}(\tfrac{1}{2\sqrt{2}}) = 0.4835 = 48.35\% \tag{26}$$

$$P_A + P_F = \frac{2}{\sqrt{\pi}} \int_{\frac{3}{2}\sqrt{2}}^{\infty} e^{-y^2} dy = \mathrm{erfc}\left(\tfrac{3}{2}\sqrt{2}\right) = 0.1336 = 13.36\% \tag{27}$$

Here we summed the two probabilities because in this way we have an integral over a symmetric domain of an even function, which we can write as the error function. If we had written P_B, P_D alone, we would have had a factor of $\frac{1}{2}$ in front of the same expression in terms of the error function. Let us rewrite the probabilities :

P_F	P_D	P_C	P_B	P_A
6.68%	24.17%	38.29%	24.17%	6.68%

(28)

If the instructor wants the other given percentages, he/she has to change the cuts: in particular, we want

$$0.20 = \mathrm{erfc}(c_A)\,, \qquad 0.40 = \mathrm{erf}(c_A) - \mathrm{erf}(c_B) = 0.8 - \mathrm{erf}(c_B)\,, \qquad 0.40 = \mathrm{erf}(c_B) \tag{29}$$

So that

$$c_A = 0.9062\,, \qquad c_B = 0.3708 \tag{30}$$

The cuts are then given by $c_A \cdot \sqrt{2} = 1.28155$ and $c_B \cdot \sqrt{2} = 0.5244$, such that the grades are: C between $\mu - 0.5244\sigma$ and $\mu + 0.5244\sigma$, B from $\mu + 0.5244\sigma$ to $\mu + 1.28155\sigma$, A above $\mu + 1.28155\sigma$, D from $\mu - 0.5244\sigma$ to $\mu - 1.28155\sigma$, F below $\mu - 1.28155\sigma$.

9 Boas, p. 770, problem 15.9-4

Suppose you receive an average of 4 calls per day. What is the probability that on a given day you receive no phone calls? Just one call? Exactly 4 calls?

We use the Poisson distribution

$$P_n = \frac{\mu^n}{n!} e^{-\mu} \tag{31}$$

where $\mu = 4$. The probability of receiving no calls will be $P_0 = e^{-4} = 0.018$, of receiving 1 call is $P_1 = 4e^{-4} = 0.073$ and the probability of receiving 4 calls is $P_4 = \frac{4^4}{4!} e^{-4} = 0.195$.

10 Boas, p. 770, problem 15.9-7

In a club of 500 members, what is the probability that exactly two people have birthdays on July 4?

The probability of having birthday on July 4 is $\frac{1}{365}$. The average number of birthdays per day will be $\mu = np = 500/365 = 1.3699$. The probability of having 2 birthdays on that same day is

$$P_2 = \frac{\mu^2}{2!} e^{-\mu} = 0.238 \tag{32}$$

11 Boas, p. 775, problem 15.10-4

Assuming a normal distribution, find the limits $\mu \pm h$ for a 90% confidence interval; for a 95% confidence interval; for a 99% confidence interval. What percent confidence is $\mu \pm 1.3\sigma$

We are looking for h such that the probability for the random variable to be between $\mu - h$ and $\mu + h$ is 90%, that is, we look for an h such that the area covered by the normal distribution is 0.90. The area is given by

$$\int_{\mu-h}^{\mu+h} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{2}{\sqrt{\pi}} \int_0^{h/\sigma} e^{-y^2/2} dy = \frac{2}{\sqrt{\pi}} \int_0^{h/(\sqrt{2}\sigma)} e^{-y^2} dy = \operatorname{erf}\left(\frac{h}{\sigma\sqrt{2}}\right); \quad (33)$$

- $\operatorname{erf}(h/(\sigma\sqrt{2})) = 0.90 \implies h = 1.645\sigma;$ (34)
- $\operatorname{erf}(h/(\sigma\sqrt{2})) = 0.95 \implies h = 1.960\sigma;$ (35)
- $\operatorname{erf}(h/(\sigma\sqrt{2})) = 0.99 \implies h = 2.576\sigma.$ (36)

12 Boas, p. 776, problem 15.10-9

Given the measurements

$$\begin{aligned} &x:\ 98,\ 101,\ 102,\ 100,\ 99 \\ &y: 21.2,\ 20.8,\ 18.1,\ 20.3,\ 19.6,\ 20.4,\ 19.5,\ 20.1 \end{aligned} \quad (37)$$

find the mean value and probable error of $x - y$, x/y, x^2y^3 and $y \ln x$.

First, let us find the mean values and standard deviations:

$$\bar{x} = \frac{1}{n}\sum_i x_i = 100\,, \qquad \sigma_x^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 = 2.5\,, \qquad \sigma_{mx} = \frac{1}{\sqrt{2}} = 0.707 \quad (38)$$

$$\bar{y} = \frac{1}{n}\sum_i y_i = 20\,, \qquad \sigma_y^2 = \frac{1}{n-1}\sum_i (y_i - \bar{y})^2 = 0.9086\,, \qquad \sigma_{my} = 0.337 \quad (39)$$

The probable error is defined as $r \equiv 0.6745\sigma_m$, which corresponds to a 50% confidence interval (see p. 774 of Boas). For our problem, the corresponding probable errors are given by $r_x = 0.477$ and $r_y = 0.227$.

If we calculate a function $w(x, y)$, the mean value and standard deviations of the mean will be:

$$\bar{w} = w(\bar{x}, \bar{y})\,, \qquad \sigma_{mw} = \sqrt{\left(\frac{\partial w}{\partial x}\right)^2 \sigma_{mx}^2 + \left(\frac{\partial w}{\partial y}\right)^2 \sigma_{my}^2}. \quad (40)$$

- $w = x - y$

$$\overline{x-y} = \bar{x} - \bar{y} = 80\,, \qquad r_w = \sqrt{r_x^2 + r_y^2} = 0.53 \quad (41)$$

- $w = x/y$:

$$\overline{x/y} = 5, \qquad r_w = \sqrt{(\tfrac{1}{20})^2 r_x^2 + (\tfrac{100}{400})^2 r_y^2} = 0.062 \quad (42)$$

- $w = x^2y^3$:

$$\overline{x^2y^3} = 8 \times 10^7, \qquad r_w = \sqrt{(200 \cdot 20^3)^2 r_x^2 + (3 \cdot 100^2 \cdot 20^2)^2 r_y^2} = 2.9 \times 10^6 \quad (43)$$

- $w = y \ln x$:

$$\overline{y \ln x} = 92, \qquad r_w = \sqrt{(20/100)^2 r_x^2 + (\ln 100)^2 r_y^2} = 1 \quad (44)$$