Probability Density and Cumulative Probability Functions

Some clarifications and conversions between Textbook by Figliola and Beasely (Wiley, 1995) and MathCAD software ***(by Prof. M. Kostic, 1997)***

There are many statistical distribution functions. The most common are ***Gaussian-normal*** distribution and ***t-Student*** distribution.

Probability Density functions

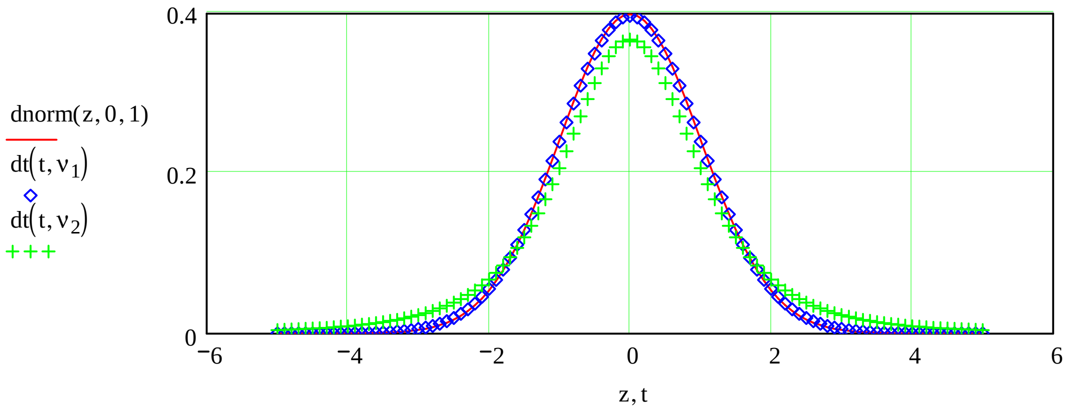
They represent the rate of probability with respect to the deviation range. Let plot the MathCAD's built-in functions, **dnorm(x, x_{mean},S_x)** for ***Gaussian-normal*** distribution, and **dt(t,n)** for ***t-Student*** distribution, please see on-line MathCAD help (if the cursor is on the function, pressing F1 key will give you the context-sensitive help).

$L := -5 \quad U := -L \quad \delta := \frac{U}{50}$ ...lower (**L**) and upper (**U**) limits, and step value (d) for independent variables **z=t=x** (normalized for **x_{mean}=0** and **S_x**=1, the **z=t"=x**)

$z := L, L+\delta .. U$

$t := L, L+\delta .. U \quad \nu_1 := 55$ ...if n is large the **dt(t,n)** will approach **dnorm(x, x_{mean},S_x)** function (see Fig. below)

$\nu_2 := 3$ ...take one small value for comparision



Cumulative or Range Probability Distribution

These are areas below the corresponding Probability-Density functions across the specified deviation range. That's where confusion begins. In different books and softwares the terminology and reference values may be different. For example, some give cumulative distribution for a range from **0** to **d**, some for **-d** to **+d**, and some others for a rage from **negative_infinity** to **d**. All express or tabulate the probability as function of single deviation value **d** or vice versa (for a reversed problem). Conversions between these are rather straightforward knowing that the total area (over the whole deviation range, -infinity to +infinity) is **1** or **100%,** see the Examples below:

The above **d=x-x_{mean}** value represents the dimensional deviation limit of any measured variable **x** from its mean value **x_{mean}**, for which the cumulative probability **P_g** or **P_{gz}** for Gaussian, or **P_t** for t-Student distribution are given in Table 4.3 (p.135) and 4.4 (p.139), respectively. The dimension**-al** deviation limit **d** is often represented as dimension**-less** value **z**(x)**=t**(x)**=d/S_x**=**(x-x_{mean})/S_x**, or some other symbol. Note, for **x_m=x_{mean}=0** and **S_x**=1, the **z=t=x**.

The cumulative or range probabilities for any deviation range (including asymmetrical or one-sided) could be determined from the available functions or tables with appropriate adjustments (see below). Sometimes a probability is given or asked for, for a variable **NOT to be** in a specified deviation range. Such probability is a complement to 100% to the "**within-the-range**" probability.

The MathCAD has several built-in cumulative probability functions, please see on-line MathCAD help (if the cursor is on the function, pressing F1 key will give you the context-sensitive help).:

$$\mathrm{pnorm}(x, x_{mean}, S_x) = \int_{-\infty}^{x} \mathrm{dnorm}(r, x_{mean}, S_x)\, dr \qquad \mathrm{cnorm}(z(x)) = \int_{-\infty}^{z} \mathrm{dnorm}(r, 0, 1)\, dr$$

$$\mathrm{pt}(t(x), \nu) = \int_{-\infty}^{t} \mathrm{dt}(r, \nu)\, dr$$ ...**r** is dummy integration variable

In our Textbook, the nomenclature is different:

$p(z(x)) = \mathrm{dnorm}(x, 0, 1)$ ..density probability function

...and then cumulative probability functions are defined for different ranges than above, i.e.::

$$P_{gz}(z) = \int_{0}^{z} p(r)\, dr$$ and $$P_t(t(x), \nu) = \int_{-t}^{t} \mathrm{dt}(r, \nu)\, dr$$

The relationships between our Textbook's function and MathCAD functions are given below:

Gaussian-Normal Distribution

(Using nomenclature from the Textbook; compare with Table 4.3, p.135):

$\mathbf{P_g}$ or $\mathbf{P_{gz}}$=Gaussian probability for **x** to be within **0** to **d**, or **0** to **z** deviation interval, respectively.

...in dimensional form

$P_g(x, x_{mean}, S_x) := \mathrm{pnorm}(x, x_{mean}, S_x) - 0.5$ $\qquad P_g(1.34, 0, 1) = 0.41$

$x(P_g, x\_mean, S_x) := \mathrm{qnorm}(P_g + 0.5, x\_mean, S_x)$ $\qquad x(0.41, 0, 1) = 1.341$

...in dimension-less form **z**(x)=**t**(x)=**d/S**$_x$=**(x-x$_{mean}$)/S**$_x$:

Remembered these!

$P_{gz}(z) := \mathrm{cnorm}(z) - 0.5$ $\qquad P_{gz}(1.34) = 0.41$ $\qquad 2 \cdot P_{gz}(1) = 68.269\,\%$

$z(P_{gz}) := \mathrm{qnorm}(P_{gz} + 0.5, 0, 1)$ $\qquad z(0.41) = 1.341$ $\qquad 2 \cdot P_{gz}(2) = 95.45\,\%$

$2 \cdot P_{gz}(3) = 99.73\,\%$

Student t-Distribution

(Using nomenclature from the Textbook; compare with Table 4.4, p.139):

$\nu = N - 1$ ...degree of freedom (**N** is number of measurements)

$t(P_t, \nu) := \mathrm{qt}\left(\frac{P_t}{2} + \frac{1}{2}, \nu\right)$ ...**t=d/S**$_x$, dim-less deviation (**d=x-x$_{mean}$**) limit, where, **S**$_x$=standard error $\qquad t(95 \cdot \%, 15) = 2.131$

$P_t(t, \nu) := 2 \cdot (\mathrm{pt}(t, \nu) - 0.5)$ ...probability for **x** to be within **-t** to **+t** deviation interval $\qquad P_t(2.131, 15) = 95\,\%$

Probability for Arbitrary Deviation Interval

(may be asymmetrical or one-sided)

If we have to calculate probability for a variable **x** to be within the values of $\mathbf{x_1}$ and $\mathbf{x_2}$, i.e. within the deviation interval $\mathbf{d_{12}=x_2-x_1}$, with number of measurements **N**, mean value $\mathbf{x_{mean}}$, and standard error $\mathbf{S_x}$ known, then a procedure may be as follows:

Given: $x_1 := 21$ $x_2 := 25$ $x_m := 20$ $S_x := 1.5$ $N := 53$ ...known (arbitrary) values

Please, enter your values for x_1 and x_2 and others above. If z_1 and z_2, or t_1 and t_2 are given (known), then enter them in place of x_1 and x_2 respectively, but then do not forget to enter values for x_m=0 and S_x=1.

Then, $z_1 := \frac{x_1 - x_m}{S_x}$ $z_2 := \frac{x_2 - x_m}{S_x}$ $t_1 := z_1$ $t_2 := z_2$ $t_1 = 0.667$ $t_2 = 3.333$

$P_{gz12} := P_{gz}(z_2) - P_{gz}(z_1)$ $P_{gz12} = 25.21\%$ ...the interval probability

Also, using dimensional values:

$P_{g1} := P_g(x_1, x_m, S_x)$ $P_{g2} := P_g(x_2, x_m, S_x)$ $P_{g1} = 0.248$ $P_{g2} = 0.5$

$P_{g12} := P_{g2} - P_{g1}$ $P_{g12} = 25.21\%$ ...must be the same as above!

Or, for the t-Student distribution

$P_{t1} := P_t(t_1, N-1)$ $P_{t2} := P_t(t_2, N-1)$ ...these are for **-t** to **+t** interval, therefore:

$P_{t12} := \frac{P_{t2}}{2} - \frac{P_{t1}}{2}$ $P_{t12} = 25.32\%$ ...is close to the above Gaussian probability if the number of measurements **N** is large.

NOTE 1:
For n=**N-1** large, $\mathbf{P_t(n,t)=2*P_{gz}(z)}$, i.e. the **t**-Student distribution approaches Gaussian-normal distribution: $P_t(1, 99) = 0.68$ and $2 \cdot P_{gz}(1) = 0.683$

The Inverse Problem

(Using nomenclature from the Textbook; compare with Table 4.3, p.135 and Table 4.4, p.139):
If the probabilities $\mathbf{P_g}$ or $\mathbf{P_{gz}}$ for Gaussian, or $\mathbf{P_t}$ for t-Student distribution are given, and the corresponding deviation intervals (**x, z,** or **t**) are to be determined, the procedure becoms inverse to the above. Then, MathCAD has handy inverse bilt-in functions with nomenclature already explained above:

...in dimensional form $x(P_g, x\_mean, S_x) := qnorm(P_g + 0.5, x\_mean, S_x)$ $x(41 \cdot \%, 0, 1) = 1.341$

...in dimension-less form $z(P_{gz}) := qnorm(P_{gz} + 0.5, 0, 1)$ $z(41 \cdot \%) = 1.341$

where, $\mathbf{P_g}$ or $\mathbf{P_{gz}}$=Gaussian probability for **x** to be within **0** to **d**, or **0** to **z** deviation interval, respectively.
Also,

$t(P_t, \nu) := qt\left(\frac{P_t}{2} + \frac{1}{2}, \nu\right)$ ...$\mathbf{t=d/S_x}$, dim-less deviation ($\mathbf{d=x-x_{mean}}$) limit, where, $\mathbf{S_x}$=standard error $t(95 \cdot \%, 15) = 2.131$

where, $\mathbf{P_t}$ is probability for t-Student distribution for dimensionles deviation to be within **-t** to **+t**, and

$\nu = N - 1$ ...degree of freedom (**N** being the number of measurements) for Student t-Distribution.