**SC531 – Lecture #9**

**STANDARD DEVIATION**

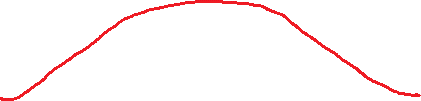
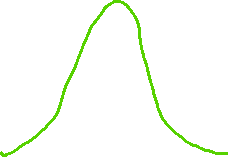
Recall our aim: Estimating the “spread" of a random variable.

Recall the pdf's of two continuous random variables X1 and X2:

pdf of X1

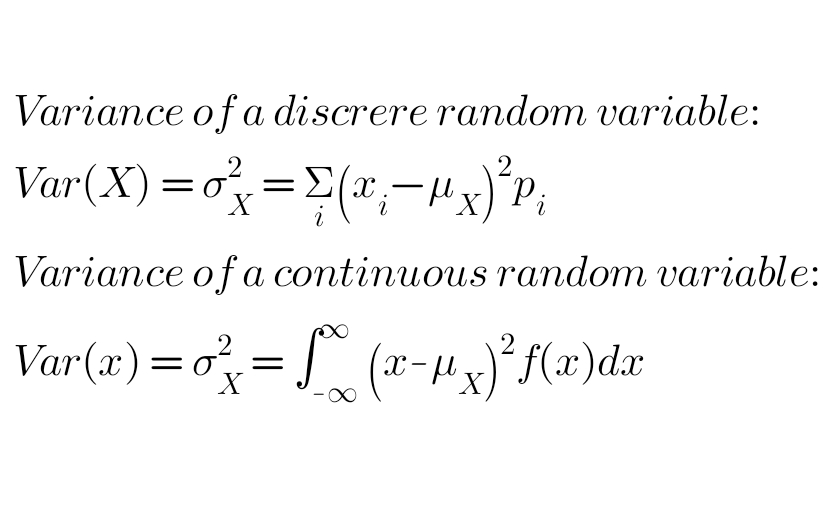
pdf of X2

X 🡪



m

The pdf's of both X1 and X2 are symmetric about their COMMON mean, shown by the blue line. Which RV has the greater spread? The answer is obvious – and it brought us to this definition:

Definition:

QUESTION: Why did we square the differences (xi – mX) and (x – mX)?

Because, in order to calculate the “spread", we chose to treat positive and negative deviations from mean m in the same way.

But that creates a huge practical difficulty!

The units of RV X and its variance s2X are not the same!

Therefore the square root of the variance of immense practical value. We call it **STANDARD DEVIATION** and denote it as sX.



**Simple & TYPICAL example:**

**The daily maximum temperature in October in a certain city follows normal distribution with mean 30oC and standard deviation 2oC. What is the probability that it will be below 26oC OR above 34oC in October next year?**

**Solution:**

**We are looking at deviations greater than 2s on either side of m.**

**Notation: The distribution of this RV, say T, is denoted by N(30,2).**

**We can standardize T by (a) subtracting mean m, and (b) dividing by standard deviation s. Denote the standardized RV as Tstd. When we do this to the two limits specified, we get, respectively, -2 and +2.**

**In other words, Tstd = (T-30)/2 🡪 Hence the new limits on Tstd.**

**Key observation: Prob( 26 < T < 34 ) = Prob( -2 < Tstd < 2 ).**

**We will see below the justification for this observation.**

**For the present, we simply refer to the tables and get the answer!**

**Note that, for N(0,1), we have F(2) = 0.9772 [see chart].**

**What is F(-2)?**

**It can easily be shown that, for N(0,1), F(-k) = 1 – F(k). This follows from the symmetry of pdf f(x).**

**So F(-2) = 0.0228.**

**So Prob( -2 < Tstd < 2 ) = 0.9772 – 0.0228 = 0.9544**

**So Prob( Tstd < -2 OR Tstd > 2 ) = 1 – 0.9544 = 0.0456**

**This is the answer to the temperature question posed above.**

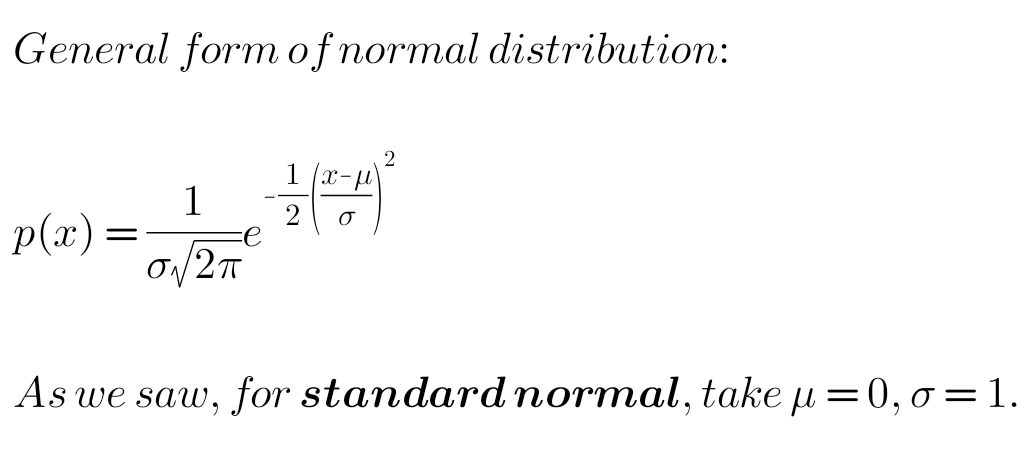
**Exercise: Similar, but with different temperature bands, such as for example 28 < T < 32, and 24 < T < 36.**

**Justification:** Suppose random variable X has mean m and standard deviation s. Then it can be shown easily that:

(1) Random variable Y defined as X - C has mean m - C and standard deviation s.

(2) Random variable Z defined as K\*X has mean K\*m and standard deviation K\*s.

In the above example, we subtracted 30 from T and then divided the result by 2, to obtain Tstd, which is N(0,1).



Further examples with calculations and charts will be worked out with the relevant spreadsheets.