# SCM Lecture 6 - teaching notes

Jeanne Wilson

September 14, 2010

## 1 Characteristics of PDFs

### 1.1 Expectation Value

This is the “centre of gravity” of the probability distribution. It’s analogous to the mean of a set of measurements. Note we can’t express the mean because these are probabilities, not values we measured, but it still tells you the most likely value if you did make the measurement.

$$E[X] = \mu = \sum_{i=1}^{N_{evt}} x_i P(x_i)$$

*Example on the board* Simple histogram with 4 values and probabilities:
P(1) = 0.2, P(2) = 0.3, P(3) = 0.4, P(4) = 0.1
$E[X] = 0.2 \times 1 + 0.3 \times 2 + 0.4 \times 3 + 0.1 \times 4 = 2.4$

### 1.2 Variance

Describes the spread of a PDF, analagous to the RMS.

$$\text{var}[X] = \sigma^2[X] = E\left[(X - E(X))^2\right]$$
$$\text{var}[X] = \sigma^2[X] = E\left[(X - \mu)^2\right]$$

We can do some manipulation to this expression. Note I have called the expectation value $\mu$ to make things clearer - this is a single value for the distribution, for a given PDF it is a constant. The expectation operator only applies to X as the expected value of a constant is the constant.

$$\begin{aligned}
\text{var}[X] &= E\left[(X - \mu)^2\right] \\
&= E\left[X^2 + \mu^2 - 2X.\mu\right] \\
&= E\left[X^2\right] + \mu^2 - 2\mu E[X] \\
&= E\left[X^2\right] + \mu^2 - 2\mu^2 \\
&= E\left[X^2\right] - \mu^2 \\
&= E\left[X^2\right] - E^2[X]
\end{aligned}$$

So lets calculate the variance of our example distribution. We already have $E[X]$ so lets calculate $E[X^2]$

$$E[X^2] = \sum_{i=1}^{N_{evt}} x_i^2 P(x_i)$$

$$E[X^2] = 0.2 \times 1 + 0.3 \times 4 + 0.4 \times 9 + 0.1 \times 16 = 6.6$$
$$\mathrm{var}[X] = 6.6 - 2.4^2 = 6.6 - 5.76 = 0.84$$

### 1.3 RMS

In the same way that we can redefine the expectation value, we can do the same manipulation to the RMS:

$$\begin{aligned}
RMS &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \\
RMS &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i^2 + \bar{x}^2 - 2x_i\bar{x})} \\
RMS &= \sqrt{(\bar{x^2} + \bar{x}^2 - 2\bar{x}^2)} \\
RMS &= \sqrt{(\bar{x^2} - \bar{x}^2)}
\end{aligned}$$

$$\bar{x^2} = \frac{1}{N}\sum_{i=1}^{N}(x^2) \quad , \quad\quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N}(x)$$

Note this is a very useful way to calculate the RMS - you can keep a running total of the sum of $x$, or $x^2$ as you go and divide by $N$ at the end.

## 2 Binomial Distribution

The binomial distribution is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments.
OK - lets take a classic example.
What is the probability of throwing two tails?
$p = \frac{1}{2}$ = the probability of throwing a tail in one toss .
Lets define our event as getting 2 tails out of the three tosses. We could have :

| | |
|---|---|
| T T H | P = pp(1-p) |
| T H T | P = p(1-p)p |
| H T T | P = (1-p)pp |

So in total $P = 3p^2(1-p)$
The formal expression for binomial probability of $n$ successes out of $N$ trials is

$$P_{N,p}(n) = \frac{N!}{n!(N-n)!}p^n(1-p)^{N-n} \quad (1)$$

where $p$ is the probability of success.
Now the binomial is a probability distribution so it won't surprise you to know that:

$$\sum_{n=0}^{N} P_{N,p}(n) = 1 \tag{2}$$

because the distribution must be normalised - the sum of all possible outcomes is 1. We can check this out for our example, N=3, for no tails n = 0,1,2,3.

$$\begin{aligned}
P_{3,p}(0) = \frac{3.2.1}{(1)(3.2.1)}p^0(1-p)^3 &= (1-p)^3 \\
P_{3,p}(1) = \frac{3.2.1}{(1)(2.1)}p^1(1-p)^2 &= 3p(1-p)^2 \\
P_{3,p}(2) = \frac{3.2.1}{(2.1)1)}p^2(1-p)^1 &= 3p^2(1-p) \\
P_{3,p}(3) = \frac{3.2.1}{(3.2.1)(1)}p^3(1-p)^0 &= p^3
\end{aligned} \tag{3}$$

Sum these together

$$\sum_{n=0}^{N} P_{N,p}(n) = (1-p)^3 + 3p(1-p)^2 + 3p^2(1-p) + p^3$$

Maybe I'll leave that to you to convince yourself it is 1 algebraically. If we use $p = 0.5$ it comes to

$$\begin{aligned}
\sum_{n=0}^{N} P_{N,p}(n) &= (0.5)^3 + \mathbf{3} \times 0.5(0.5)^2 + \mathbf{3} \times 0.5^2(0.5) + 0.5^3 \\
\sum_{n=0}^{N} P_{N,p}(n) &= 8 \times (0.5)^3 = 8 \times 0.125 = 1
\end{aligned}$$

Note the coefficient that come from the factorials - in this case, 1,3,3,1 are the number of outcomes - we can only get no tails in one way - HHH and all tails - TTT but there are three ways of getting 2T or 1T. Note also - given that there are only two outcomes these coefficients are always symmetric. I'm sure you've all seen Pascal's triangle before:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
...

What if our coin was biased and $p = 0.4$? Lets calculate, again for $N = 3$.

$$\begin{aligned}
P_{3,p}(0) &= (1-p)^3 = 0.6^3 = 0.216 \\
P_{3,p}(1) &= 3p(1-p)^2 = 3 \times 0.4 \times 0.6^2 = 0.432 \\
P_{3,p}(2) &= 3p^2(1-p) = 3 \times 0.4^2 \times 0.6 = 0.288 \\
P_{3,p}(3) &= p^3 = 0.4^3 = 0.064
\end{aligned} \tag{4}$$

Now for small $N$ the *probability* distribution is not symmetric even though the coefficients are. However, as N increases, the distribution becomes more symmetric, even for values of $p$ far from 0.5.

*Show some distributions*

### 2.1 Mean

The mean for the binomial distribution is the mean number of successes:

$$E[n] = \sum_{n=0}^{N} n.P_{N,p} = Np \tag{5}$$

### 2.2 Variance

Again this is the spread of the binomial distribution

$$\sigma_n^2 = \mathrm{var}[n] = Np(1-p) \tag{6}$$

## 3 An Example

Opinion polls in a 2 horse race. Obama or McCain? 1000 randomly selected people were polled - ie $N = 1000$. 510 people voted for Obama and 490 for McCain. Lets define
$p =$ probability Obama get the vote $= \frac{510}{1000} = 0.51$
So we can say that Obama got 51% of the vote but what is the uncertainty on that? Use the square root of the variance: $\sigma_n = \sqrt{Np(1-p)} = \sqrt{1000 \times 0.51 \times 0.49} = \sqrt{249.9} = 15.8$ people.
So what does that mean in terms of significance? Its easier if we consider the fractional uncertainty $= \frac{\sigma_n}{Np} = \frac{15.8}{510} \approx 3\%$ So our poll result shows no significant lead: Obama has $51 \pm 3\%$ of the vote.
How many people would we have to have polled to see a swing of 1%?
We want $\pm\frac{\sigma_n}{Np} = 0.01 = \frac{\sqrt{Np(1-p)}}{Np} = \sqrt{\frac{Np(1-p)}{N^2p^2}} = \sqrt{\frac{\frac{1}{p}-1}{N}}$
ie. $N = \frac{\frac{1}{p}-1}{0.01^2} = \frac{\frac{1}{0.51}-1}{0.01^2} = 9607$ so we need to poll more than 9 times the amount of people to get $\pm 1\%$ accuracy in the poll.

(7)