CFCS
Expectation and Variance; Chebyshev’s Theorem

Miles Osborne (originally: Frank Keller)

School of Informatics
University of Edinburgh
miles@inf.ed.ac.uk

February 8, 2008
1 Expectation and Related Concepts
   • Expectation
   • Mean
   • Variance

2 Chebyshev’s Theorem
Much of probability theory comes from gambling. If we bought a lottery ticket, how much would we expect to win on average?

**Example**

In a raffle, there are 10,000 tickets. The probability of winning is therefore \( \frac{1}{10,000} \) for each ticket. The prize is worth $4,800. Hence the *expectation* per ticket is \( \frac{4,800}{10,000} = 0.48 \).

In this example, the expectation can be thought of as the average win per ticket.
This intuition can be formalized as the expected value of a random variable:

**Definition: Expected Value**

If $X$ is a discrete random variable and $f(x)$ is the value of its probability distribution at $x$, then the expected value of $X$ is:

$$E(X) = \sum_{x} x \cdot f(x)$$

We will only deal with the discrete case here (but the definition can be extended to cover continuous random variables).
A balanced coin is flipped three times. Let $X$ be the number of heads. Then the probability distribution of $X$ is:

$$f(x) = \begin{cases} \frac{1}{8} & \text{for } x = 0 \\ \frac{3}{8} & \text{for } x = 1 \\ \frac{3}{8} & \text{for } x = 2 \\ \frac{1}{8} & \text{for } x = 3 \end{cases}$$

The expected value of $X$ is:

$$E(X) = \sum_{x} x \cdot f(x) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}$$
The notion of expectation can be generalized to cases in which a function $g(X)$ is applied to a random variable $X$.

**Theorem: Expected Value of a Function**

If $X$ is a discrete random variable and $f(x)$ is the value of its probability distribution at $x$, then the expected value of $g(X)$ is:

$$E[g(X)] = \sum_x g(x)f(x)$$
Example

Let $X$ be the number of points rolled with a balanced die. Find the expected value of $X$ and of $g(X) = 2X^2 + 1$.

The probability distribution for $X$ is $f(x) = \frac{1}{6}$. Therefore:

$$E(X) = \sum_{x} x \cdot f(x) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{21}{6}$$

$$E[g(X)] = \sum_{x} g(x) f(x) = \sum_{x=1}^{6} (2x^2 + 1) \frac{1}{6} = \frac{94}{6}$$
The expectation of a random variable is also called the _mean_ of the random variable. It’s denoted by $\mu$.

**Definition: Mean**

If $X$ is a discrete random variable and $f(x)$ is the value of its probability distribution at $x$, then the mean of $X$ is:

$$
\mu = E(X) = \sum_x x \cdot f(x)
$$

Intuitively, $\mu$ denotes the _average_ value of $X$. 
Histogram with mean for the distribution in the previous example (number of heads in three coin flips):

\[ E(X) = \mu \]
Definition: Variance

If $X$ is a discrete random variable and $f(x)$ is the value of its probability distribution at $x$, and $\mu$ is its mean then:

$$\sigma^2 = \text{var}(X) = E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x)$$

is the variance of $X$.

Intuitively, $\text{var}(X)$ reflects the *spread* or *dispersion* of a distribution, i.e., how much it diverges from the mean.

$\sigma$ is called the *standard deviation* of $X$. 

Miles Osborne (originally: Frank Keller)
Example

Let $X$ be a discrete random variable with the distribution:

$$f(x) = \begin{cases} 
\frac{1}{8} & \text{for } x = 0 \\
\frac{3}{8} & \text{for } x = 1 \\
\frac{3}{8} & \text{for } x = 2 \\
\frac{1}{8} & \text{for } x = 3 
\end{cases}$$

Then the variance and standard deviation of $X$ are:

$$\text{var}(X) = \sum_x (x - \mu)^2 f(x)$$

$$= (0 - \frac{3}{2})^2 \cdot \frac{1}{8} + (1 - \frac{3}{2})^2 \cdot \frac{3}{8} + (2 - \frac{3}{2})^2 \cdot \frac{3}{8} + (3 - \frac{3}{2})^2 \cdot \frac{1}{8}$$

$$= 0.86$$

$$\sigma = \sqrt{\text{var}(X)} = 0.93$$
Variance

Histogram with mean and standard deviation for the previous example:
\[ \sigma^2 \text{ as a measure of dispersion:} \]

\[ \mu = 5 \text{ and } \sigma^2 = 5.26 \]

\[ \mu = 5 \text{ and } \sigma^2 = 3.18 \]
\( \sigma^2 \) as a measure of dispersion:

\[
\begin{align*}
\mu &= 5 \quad \text{and} \quad \sigma^2 = 1.66 \\
\mu &= 5 \quad \text{and} \quad \sigma^2 = 0.88
\end{align*}
\]
Chebyshev’s Theorem

Chebyshev’s Theorem

If $\mu$ and $\sigma$ are the mean and the standard deviation of a random variable $X$, and $\sigma \neq 0$, then for any positive constant $k$:

$$P(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

In other words, the probability that $X$ will take on a value within $k$ standard deviations of the mean is at least $1 - \frac{1}{k^2}$.

Example

Assume $k = 2$. Then $P(|x - \mu| < 2\sigma) = 1 - \frac{1}{2^2} = \frac{3}{4}$, i.e., at least 75% of the values of $X$ fall within 2 standard deviations of the mean.
Chebyshev’s Theorem

Example: distribution with $\mu = 4.99$ and $\sigma = 3.13$. 
Chebyshev’s Theorem

Example

Using Chebyshev’s Theorem, we can show: if $X$ is normally distributed, then:

$$P(|x - \mu| < 2\sigma) = .9544$$

In other words, the 95.44% of all values of $X$ fall within 2 standard deviations of the mean. This is a tighter than the bound of 75% that holds for an arbitrary distribution.

Many cognitive variables (e.g., IQ measurements) are normally distributed. More on this in the next lecture.
Example: normal distribution with $\mu = 0$ and $\sigma = 1$. 
Summary

- The expected value of a random variable is its average value over a distribution;
- the mean is the same as the expected value;
- the variance of random variable indicates its dispersion, or spread around the mean;
- Chebyshev’s theorem places a bound on the probability that the values of a distribution will be within a certain interval around the mean;
- for example, at least 75% of all values of a distribution fall within two standard deviations of the mean.