# CMPSCI 240: "Reasoning Under Uncertainty"

## Lecture 14

Prof. Hanna Wallach
wallach@cs.umass.edu

March 8, 2012

# Reminders

- Check the course website: `http://www.cs.umass.edu/~wallach/courses/s12/cmpsci240/`
- Fifth homework is due TOMORROW

# Recap

# Last Time: The Markov Inequality

- Markov inequality: for any nonnegative random variable $X$

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \quad \text{for all } a > 0$$

# Last Time: The Markov Inequality

- Markov inequality: for any nonnegative random variable $X$

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \quad \text{for all } a > 0$$

- Intuition: if a nonnegative random variable $X$ has a small mean, then the probability that $X$ is large is small

# Last Time: The Chebyshev Inequality

- Chebyshev inequality: For any random variable $X$

$$P(|X - \mathbb{E}[X]| \geq c) \leq \frac{\text{var}(X)}{c^2} \quad \text{for } c > 0$$

# Last Time: The Chebyshev Inequality

- Chebyshev inequality: For any random variable $X$

$$P(|X - \mathbb{E}[X]| \geq c) \leq \frac{\text{var}(X)}{c^2} \quad \text{for } c > 0$$

- Intuition: if random variable $X$ has a small variance, then the probability that the value of $X$ is far from its mean is small

# Last Time: The Chebyshev Inequality

- Chebyshev inequality: For any random variable $X$

$$P(|X - \mathbb{E}[X]| \geq c) \leq \frac{\text{var}(X)}{c^2} \quad \text{for } c > 0$$

- Intuition: if random variable $X$ has a small variance, then the probability that the value of $X$ is far from its mean is small
- Note that $X$ does not need to be nonnegative

# Last Time: Limitations of the Markov Inequality

- Bounds provided by the Markov inequality can be quite loose
- Can obtain tighter bounds if we use the variance as well

- Markov is meaningless when $\mathbb{E}[X] > a$
- Chebyshev is meaningless when $\texttt{var}(X) > c^2$

# The Weak Law of Large Numbers

# The Weak Law of Large Numbers

- If $X_1, X_2, \ldots, X_n$ are independent identically-distributed random variables with mean $\mathbb{E}[X_i]$, then for every $\epsilon > 0$

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}[X_i]\right| \geq \epsilon\right) \to 0 \quad \text{as } n \to \infty$$

- Intuition: the sample mean of a very large number of independent identically distributed random variables is very close to the true mean with high probability

# Examples of the Weak Law of Large Numbers

- e.g., suppose we would like to estimate the president's approval rating $p$. We ask $n$ random voters whether or not they approve of the president and use the fraction of voters who approve as our estimate. If we would like to have high confidence (e.g., 95%) that our estimate is very accurate (i.e., within 0.01 of the true approval rating) how many voters should we poll? Hint: suppose $p(1-p) \leq 1/4$

# For Next Time

- Check the course website: `http://www.cs.umass.edu/~wallach/courses/s12/cmpsci240/`
- Fifth homework is due TOMORROW