CS70 In Simpler Terms - Note 8

written by Kevin Liu cs70.kevinmliu.com

July 29, 2017

1 Markov's and Chebyshev's Inequality

The following Inequalities are all interrelated and are used to bound the probability of deviating from the mean.

- Markov's Inequality: $Pr(X \geqslant a) \leqslant \frac{E(f(x))}{f(a)}$ where f is an increasing, non-negative function
- Weak form of Markov's Inequality (the form you will most often use): $Pr(X \geqslant a) \leqslant \frac{E(X)}{a}$ where X is non-negative
- Chebyshev's Inequality is a continuation of Markov's Inequality. It attempts to bound the probability of any deviation α above or below the mean:
  - $Pr(|X - \mu| \geqslant \alpha) \leqslant \frac{Var(X)}{\alpha^2}$
  - The result makes sense because if the $Var(X)$ decreases, then the probability of larger deviations also decreases
  - If we let $\alpha = \beta\sigma$, we can achieve the probability of getting within β standard deviations from the mean: $Pr(|X - \mu| \geqslant \beta\sigma) \leqslant \frac{1}{\beta^2}$
  - Chebyshev's Inequality can be applied to *any* r.v.

2 Confidence Interval

- Confidence is defined as 1 - δ, so a δ of 0.05 gives a confidence of 0.95
- $Pr[\mu \in (\bar{x} - \epsilon, \bar{x} + \epsilon)] = Pr(|\bar{x} - \mu| < \epsilon)$
  $=$ 1 - $Pr(|\bar{X} - \mu| \geqslant \alpha)$ $(Chebyshev)$
  $\geqslant 1 - \frac{Var(\bar{X})}{\epsilon^2} = 1 - \frac{\sigma^2}{n\epsilon^2} \geqslant 1 - \delta$
- $1 - \frac{\sigma^2}{n\epsilon^2} \geqslant 1 - \delta \Rightarrow \delta \geqslant \frac{\sigma^2}{n\epsilon^2}$
  (at most δ error and at least 1 - δ confident)

An important result to keep in mind is that as n increases, the Pr(specific event) decreases, but the Pr(near the event) increases.
Example:
Oftentimes you will be asked to find n given some amount of confidence. We will explore that in this example.

- We attempt to estimate the bias of a coin (e.g. 0.5 for a fair coin)
- Let S_n = number of heads in n tosses, and let $\hat{p}$ = bias estimate
- $\hat{p} = \frac{S_n}{n}$, $S_n = \sum_{i=1}^{n} X_i \begin{cases} X_i = 1, heads \\ X_i = 0, tails \end{cases}$
- $E(X_i) = p, E(\hat{p} = E(\frac{S_n}{n}) = \frac{1}{n} n X_i = p$
- $Var(\hat{p}) = Var(\frac{S_n}{n} = \frac{1}{n^2} Var(S_n)) = \frac{\sigma^2}{n}$
- $Pr(|\hat{p} - p| \geqslant \epsilon) \leqslant \frac{\sigma^2}{n\epsilon^2}$
  $\Rightarrow n \geqslant \frac{\sigma^2}{\epsilon^2 \delta} = \frac{p(1-p)}{\epsilon^2 \delta} \geqslant \frac{1}{4\epsilon^2 \delta}$ (variance of single sample is $p(1-p)$ and the bound is at $p = 0.5$)

3 Weak Law of Large Numbers

Given an n sample test:

- Law of Large Numbers - more samples $\Rightarrow$ less deviations
- $\bar{X} = \frac{X_1 + X_2 + ... X_n}{n}$
- $E(\bar{X}) = \frac{E(X_1) + ... E(X_n)}{n} = \frac{n\mu}{n} = \mu$
- $Var(\bar{X}) = Var(\frac{1}{n}(X_1 + X_2 + ... X_n)) = \frac{1}{n^2}(Var(X_1) + ... Var(X_n)) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$
- $Var(\bar{X}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$ and $\bar{X}$ converges to μ