Markov and Chebyshev Inequalities

Dirk Grunwald
University of Colorado, Boulder

- How much do we have to measure things to have a probabilistic bound on their frequency?
- Assume a simple Bernoulli trial for an event A (e.g. coin is heads).
- If S_n is the sum of true events, we expect $E[A] = S_n/n$.

- How much do we have to measure things to have a probabilistic bound on their frequency?
- Assume a simple Bernoulli trial for an event A (e.g. coin is heads).
- If S_n is the sum of true events, we expect $E[A] = S_n/n$.
- Clearly not likely when $n = 1$.

Limits & Bounds

- How much do we have to measure things to have a probabilistic bound on their frequency?
- Assume a simple Bernoulli trial for an event A (e.g. coin is heads).
- If S_n is the sum of true events, we expect $E[A] = S_n/n$.
- Clearly not likely when $n = 1$.
- How large does n need to be to get a reasonable estimate of $E[A]$?

Goal: Weak Law of Large Number

- The weak law of large numbers defines how much we must measure to estimate the mean with a specific accuracy.
- The weak law is simple to prove and helps us understand a stronger method called the Central Limit Theorem.
- Valid when $E[X]$ is finite (e.g. not Cauchy or some Pareto)

Reading Inequalities

- The Weak Law builds on two inequalities or bounds on the probability of something happening: the Markov Inequality and the Chebyshev Inequality.
- Those inequalities estimate properties about a distribution when we only know trivial properties.

- Markov Inequality

$$P[X \geq tE[X]] \leq 1/t$$

- a very imprecise inequality when you only know the mean

- Markov Inequality

$$P[X \geq tE[X]] \leq 1/t$$

- a very imprecise inequality when you only know the mean

- Chebyshev's Inequality - $P[|X - E[x]| \geq t] \leq \frac{\sigma^2}{t^2}$ - when you know mean and variance, the probability of samples falling more than t from the mean is related to the variance and t.

Markov Inequailty

$$E[X] = \sum_{x_i} x_i p(x_i)$$

Markov Inequailty

$$
\begin{aligned}
E[X] &= \sum_{x_i} x_i p(x_i) \\
&= \sum_{i<t} x_i p(x_i) + \sum_{t \leq x_i} x_i p(x_i)
\end{aligned}
$$

Markov Inequailty

$$
\begin{aligned}
E[X] &= \sum_{x_i} x_i p(x_i) \\
&= \sum_{i<t} x_i p(x_i) + \sum_{t \le x_i} x_i p(x_i) \\
&\ge \sum_{t \le x_i} x_i p(x_i) \qquad \text{(discard left half)}
\end{aligned}
$$

Markov Inequailty

$$\begin{aligned}
E[X] &= \sum_{x_i} x_i p(x_i) \\
&= \sum_{i<t} x_i p(x_i) + \sum_{t\le x_i} x_i p(x_i) \\
&\ge \sum_{t\le x_i} x_i p(x_i) \qquad \text{(discard left half)} \\
&\ge \sum_{t\le x_i} t p(x_i) \qquad \text{(all } x_i \ge t\text{)}
\end{aligned}$$

Markov Inequailty

$$
\begin{aligned}
E[X] &= \sum_{x_i} x_i p(x_i) \\
&= \sum_{i<t} x_i p(x_i) + \sum_{t \leq x_i} x_i p(x_i) \\
&\geq \sum_{t \leq x_i} x_i p(x_i) && \text{(discard left half)} \\
&\geq \sum_{t \leq x_i} t p(x_i) && \text{(all } x_i \geq t\text{)} \\
&\geq t P[X \geq t] && \text{(def. of } P[X \geq t]\text{)}
\end{aligned}
$$

Markov Inequailty

$$\begin{aligned}
E[X] &= \sum_{x_i} x_i p(x_i) \\
&= \sum_{i<t} x_i p(x_i) + \sum_{t \le x_i} x_i p(x_i) \\
&\ge \sum_{t \le x_i} x_i p(x_i) && \text{(discard left half)} \\
&\ge \sum_{t \le x_i} t p(x_i) && \text{(all } x_i \ge t\text{)} \\
&\ge t P[X \ge t] && \text{(def. of } P[X \ge t]\text{)} \\
\frac{E[X]}{t} &\ge P[X \ge t]
\end{aligned}$$

Markov Inequailty - Rearranged

You can re-arrange the markov inequality to get a rough estimate of something being more than t times the mean.

$$P[X \geq t] \quad \leq \quad \frac{E[X]}{t}$$

Markov Inequailty - Rearranged

You can re-arrange the markov inequality to get a rough estimate of something being more than t times the mean.

$$
\begin{aligned}
P[X \geq t] &\leq \frac{E[X]}{t} \\
P[X \geq tE[X]] &\leq \frac{E[X]}{tE[X]}
\end{aligned}
$$

Markov Inequailty - Rearranged

You can re-arrange the markov inequality to get a rough estimate of something being more than t times the mean.

$$
\begin{aligned}
P[X \geq t] &\leq \frac{E[X]}{t} \\
P[X \geq tE[X]] &\leq \frac{E[X]}{tE[X]} \\
P[X \geq tE[X]] &\leq \frac{1}{t}
\end{aligned}
$$

The probability of drawing a sample greater than twice the mean is less than 50%