Contents

Why are Inequalities Useful?

We may not know the true form of a probability distribution

- Opinion polls
- Stock market data
- Weather tomorrow

But we may know some of its priorities

- Mean
- Variance
- Non-negativity

Inequalities let us say something about the probability distibution is such cases, although often imprecise. They are also very important for looking at what happens as we collect more and more measurements.

Markov's Inequality

Often we want to know:

> What is the probability that the value of a random variable X is "far" from its mean?

A generic answer for non-negative X is Markov's inequality. Say X is a non-negative random variable. Then:

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

Andrew Andreyevish Markov (1856-1922) was a Russian Mathematician

- Markov's inequality is named after him
- Also Markov Chains used, e.g. in Google's PageRank algorithm

Example: Roll a 6 sides dice

- Mean is $E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$
- Morckov inequality: $P(X \geq 5) \leq \frac{3.5}{5} = 0.7$. Exact: $P(X = 5) + P(X = 6) = \frac{1}{6} + \frac{1}{6} = 0.33$
- So a loose bound, but is made *no* assumptons about the form of the distribution.

Example: Distribution of number X of facebook friends

- Mean if $E(X) = 190(!)$
- Markov inequality: $P(X \geq 500) \leq \frac{190}{500} = 0.38$. From plot, $P(X \geq 500) \approx 0.1$
- Markov inequality: $P(X \geq 190) \leq \frac{190}{190} = 1$, non-informative. From splot, $P(X \geq 190) \approx 0.3$

Chebyshev's Inequality

Suppose X is a random variable with mean $E(X) = \mu$ and variance $var(X) = \sigma^2$. Then

$$P(\mid X - \mu \mid \geq k) \leq \frac{\sigma^2}{k^2} \text{ for all } k > 0$$

Pafnuty Lvovish Chebyshev (1821-1894) also a Russian mathematician.

- Chebyshev's inequality was in fact first formulated by French mathematician Jules Bienaym'{e} without proof, then proved by Chebyshev 14 years later
- Markov was a graduate student of Chebyshev (also Aleksandr Lyapunov, but that's another days work)

Chebyshev's inequality links the "spread" of values of a random variable around its mean to the variance σ^2:

- Applying Chebyshev's inequality $P(\mid X - \mu \mid \geq k) \leq \frac{\sigma^2}{k^2}$ with $k = n\sigma$ gives:

$$P(\mid X - \mu \mid \geq n\sigma) \leq \frac{1}{n^2}$$

2

- With $n = 3$ then $P(| X - \mu | \geq 3\sigma) \leq \frac{1}{9} = 0.11$
- This holds even when distribution is not Gaussian, so can be quite handy (if conservative).

Example: Roll 6 sided dice

- Mean is $E[X] = 1 \times \frac{1}{6} + 2 \times frac16 + \cdots + 6 \times \frac{1}{6} = 3.5$
- Variance is $Var(X) = E[X^2] - E[X]^2$
    - $E[X^2] = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \cdots + 6^2 \times \frac{1}{6} \approx 15.17$
    - $Var(X) = 15.17 - 3.5^2 \approx 2.9$
- Chebyshev inequality: $P(| X - 3.5 | \geq 2.5) \leq \frac{2.9}{2.5^2} = 0.46$
- Exact: $P(| X - 3.5 | \geq 2.5) = P(X = 1) + P(X = 6) = \frac{1}{6}\frac{1}{6} = 0.33$
- A loose bound, but use of variance in Chebyshev inequality can improve accuracy of Markov inequality.

Example: IQ in Ireland

- Mean is $E(X) = 92$, variance is $\sigma^2 = 225$
- Chebyshev inequality: $P(| X - 92 | \geq 20) \leq frac225400 = 0.56$. From data, $P(| X - 92 | \geq 20) \approx 0.18$.
- Markov inequality: $P(X \geq 112) \leq \frac{92}{112} = 0.82$. And need to add $P(X \leq 72)$ to this.

Chernoff's Inequality

Suppose we have random variable X. Then

$$P(X \geq a) \leq \min_{t>0} e^{ta} e^{\log E(e^{tX})}$$

Herman Chernoff is a US mathematician (with Russian parents).

- Was at MIT, then Harvard

$E(e^{tX})$ is called the "moment generating function"

Contains more information about the distribution than just themean (used by Markov inequality) and variance (used by Chebyshev inequality).

Example: coin flipping

- A fair coin lands on heads with probability $\frac{1}{2}$ and on tails with probability $\frac{1}{2}$

- If a ocin is flipped 100 times, give an upper bound on the probability that it lands heads at least 60 times

- Random variable $X_k = 1$ is heads, $X + 0$ if tails at flip k.

  - $S = \sum_{k=1}^{100} K_k$

- $E(e^{tX_k}) = \frac{1}{2}e^{t\times1} + \frac{1}{2}e^{t\times0} = \frac{1}{2}(e^t + 1)$

$$\log E(e^{tS}) = \log E(\prod_{k=1}^{100} e^{tX_k}) = \log \prod_{k=1}^{100} E(e^{tX_k}) = 100 \log(\frac{1}{2}(e^t + 1))$$

- By Chernoff's inequality, the probability of at least 55 heads is

$$P(S \geq 60) \leq \min_{t>0} e^{-60t} e^{\log E(e^{tX})} = \min_{t>0} e^{-60t} e^{100 \log(\frac{1}{2}(e^t+1))}$$

Example: $P(S \geq 60) \leq min_{t>0} e^{-60t} e^{100 \log(\frac{1}{2}(e^t+1))}$

- Using $t = 0.4$, Chenoff's inequality gives $P(S \geq 60) \leq 0.13$
- Marckov inequality gives $P(S \geq 60) \leq \frac{E(S)}{60} = \frac{50}{60} = 0.83$