CMPT 210: Probability and Computing

Lecture 20

Sharan Vaswani

March 23, 2023

Tail inequalities

Variance gives us one way to measure how "spread" the distribution is.

Tail inequalities bound the probability that the r.v. takes a value much different from its mean.

Example: Consider a r.v. X that can take on only non-negative values and $\mathbb{E}[X] = 99.99$. Show that $\Pr[X \geq 300] \leq \frac{1}{3}$.

$$\begin{aligned} \textit{Proof}: \mathbb{E}[X] &= \sum_{x \in \text{Range}(X)} x \Pr[X = x] = \sum_{x | x \geq 300} x \Pr[X = x] + \sum_{x | 0 \leq x < 300} x \Pr[X = x] \\ &\geq \sum_{x | x \geq 300} (300) \Pr[X = x] + \sum_{x | 0 \leq x < 300} x \Pr[X = x] \\ &= (300) \Pr[X \geq 300] + \sum_{x | 0 \leq x < 300} x \Pr[X = x] \end{aligned}$$

If $\Pr[X \geq 300] > \frac{1}{3}$, then, $\mathbb{E}[X] > (300)\,\frac{1}{3} + \sum_{x | 0 \leq x < 300} x \Pr[X = x] > 100$ (since the second term is always non-negative). Hence, if $\Pr[X \geq 300] > \frac{1}{3}$, $\mathbb{E}[X] > 100$ which is a contradiction since $\mathbb{E}[X] = 99.99$.

Markov's Theorem

Markov's theorem formalizes the intuition on the previous slide, and can be stated as follows.

Markov's Theorem: If X is a non-negative random variable, then for all $x > 0$,

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

Proof: Define $\mathcal{I}\{X \geq x\}$ to be the indicator r.v. for the event $[X \geq x]$. Then for all values of X, $x\,\mathcal{I}\{X \geq x\} \leq X$.

$$\mathbb{E}[x\,\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\,\mathbb{E}[\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\Pr[X \geq x] \leq \mathbb{E}[X]$$
$$\implies \Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

Since the above theorem holds for all $x > 0$, we can set $x = c\mathbb{E}[X]$ for $c \geq 1$. In this case, $\Pr[X \geq c\mathbb{E}[X]] \leq \frac{1}{c}$. Hence, the probability that X is "far" from the mean in terms of the multiplicative factor c is upper-bounded by $\frac{1}{c}$.

Markov's Theorem – Example

Q: Suppose there is a dinner party where n people check in their coats. The coats are mixed up during dinner, so that afterward each person receives a random coat. In particular, each person gets their own coat with probability $\frac{1}{n}$.

Recall that if G is the r.v. corresponding to the number of people that receive their own coat, then we used the linearity of expectation to derive that $\mathbb{E}[G] = 1$. Using Markov's Theorem,

$$\Pr[G \geq x] \leq \frac{\mathbb{E}[G]}{x} = \frac{1}{x}.$$

Hence, we can bound the probability that x people receive their own coat. For example, there is no better than 20% chance that more than 5 people get their own coat.

Markov's Theorem – Example

Q: If X is a non-negative r.v. such that $\mathbb{E}[X] = 150$, compute the probability that X is at least 200. Ans: $\Pr[X \geq 200] \leq \frac{\mathbb{E}[X]}{200} = \frac{3}{4}$

Q: If we are provided additional information that X can not take values less than 100 and $\mathbb{E}[X] = 150$, compute the probability that X is at least 200.

Define $Y := X - 100$. $\mathbb{E}[Y] = \mathbb{E}[X] - 100 = 50$ and Y is non-negative.

$$\Pr[X \geq 200] = \Pr[Y + 100 \geq 200] = \Pr[Y \geq 100] \leq \frac{\mathbb{E}[Y]}{100} = \frac{50}{100} = \frac{1}{2}$$

Hence, if we have additional information (in the form of a lower-bound that a r.v. can not be smaller than some constant $b > 0$), we can use Markov's Theorem on the shifted r.v. (Y in our example) and obtain a tighter bound on the probability of deviation.

Chebyshev's Theorem

Chebyshev's Theorem: For a r.v. X and any constant $y > 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq y] \leq \frac{\mathrm{Var}[X]}{y^2}.$$

Proof: Use Markov's Theorem with some cleverly chosen function of X. Formally, for some function f such that $Y := f(X)$ is non-negative. Using Markov's Theorem for Y,

$$\Pr[f(X) \geq x] \leq \frac{\mathbb{E}[f(X)]}{x}$$

Choosing $f(X) = |X - \mathbb{E}[X]|^2$ and $x = y^2$ implies that $f(X)$ is non-negative and $x > 0$. Using Markov's Theorem,

$$\Pr[|X - \mathbb{E}[X]|^2 \geq y^2] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{y^2}$$

Note that $\Pr[|X - \mathbb{E}[X]|^2 \geq y^2] = \Pr[|X - \mathbb{E}[X]| \geq y]$, and hence,

$$\Pr[|X - \mathbb{E}[X]| \geq y] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{y^2} = \frac{\mathrm{Var}[X]}{y^2}$$

Chebyshev's Theorem

Chebyshev's Theorem bounds the probability that the random variable X is "far" away from the mean $\mathbb{E}[X]$ by an additive factor of x.

If we set $x = c\sigma_X$ where σ_X is the standard deviation of X, then by Chebyshev's Theorem,

$$\Pr[(X \geq \mathbb{E}[X] + c\,\sigma_X) \cup (X \leq \mathbb{E}[X] - c\,\sigma_X)] = \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \leq \frac{\mathrm{Var}[X]}{c^2\sigma_X^2} = \frac{1}{c^2}$$

$$\begin{aligned} &\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = \Pr[|X - \mathbb{E}[X]| \leq c\sigma_X] \\ \implies &\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = 1 - \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \geq 1 - \frac{1}{c^2}. \end{aligned}$$

Chebyshev's Theorem is used to bound the probability that X is "concentrated" near its mean.

Unlike Markov's Theorem, Chebyshev's Theorem does not require the r.v. to be non-negative, but requires knowledge of the variance.

Chebyshev's Theorem - Example

Q: If X is a non-negative r.v. such that $\mathbb{E}[X] = 100$ and $\sigma_X = 15$, compute the probability that X is at least 300.

If we use Markov's Theorem, $\Pr[X \geq 300] \leq \frac{\mathbb{E}[X]}{300} = \frac{1}{3}$.

Note that $\Pr[|X - 100| \geq 200] = \Pr[X \leq -100 \cup X \geq 300] = \Pr[X \geq 300]$. Using Chebyshev's Theorem,

$$\Pr[X \geq 300] = \Pr[|X - 100| \geq 200] \leq \frac{\mathrm{Var}[X]}{(200)^2} = \frac{15^2}{200^2} \approx \frac{1}{178}.$$

Hence, by exploiting the knowledge of the variance and using Chebyshev's inequality, we can obtain a tighter bound.