STAT 315: Markov, Chebyshev, Hoeffding and Confidence Intervals

Luc Rey-Bellet

University of Massachusetts Amherst

luc@math.umass.edu

April 15, 2025

Markov and Chebyshev Inequalities

Idea: How to extract information from the mean and the variance.

Two inequalities

- Markov: If Y is a non-negative RV with $\mu = E[Y]$ then for any $a > 0$

$$P(Y \geq a) \leq \frac{E[Y]}{a} = \frac{\mu}{a}.$$

- Chebyshev: If Y is a RV with $\mu = E[Y] = \sigma^2 = V(Y)$ then for any $\epsilon > 0$

$$P(|Y - \mu| \geq \epsilon) \leq \frac{V[Y]}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}.$$

$$\text{or } P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Proof

Let us do the proof for continuous random variable. For Markov if $Y \geq 0$ then

$$E[Y] = \int_0^\infty yf(y)dy \geq \int_a^\infty yf(y)dy$$
$$\geq \int_a^\infty af(y)dy = a\int_a^\infty f(y)dy = aP(Y \geq a)$$

Chebyshev is a consequence of Markov for the positive $(Y-\mu)^2$

$$P(|Y-\mu| \geq \epsilon) = P((Y-\mu)^2 \geq \epsilon^2)$$
$$= \frac{E[(Y-\mu)^2]}{\epsilon^2} \qquad \text{by Markov inequality}$$
$$= \frac{V(Y)}{\epsilon^2} \qquad \text{by definition of the variance}$$

If we take $\epsilon = k\sigma$ to be a multiple of the standard deviation we obtain the unit-free version $P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

Example

- Suppose Y is a Poisson random variable with parameter $\lambda = 2$.
- Recall $E[Y] = \lambda$ and $V[Y] = \lambda$.
- Estimate the probability that Y exceed 4 times its mean, that is

$$P(Y \geq 4E[Y]) = P(Y \geq 4\lambda)$$

- Using Markov inequality we have

$$P(Y \geq 4\lambda) \leq \frac{E[Y]}{4\lambda} = \frac{\lambda}{4\lambda} = \frac{1}{4}$$

- Using Chebyshev inequality we have

$$\begin{aligned} P(Y \geq 4\lambda) = P(Y - \lambda \geq 3\lambda) \leq P(|Y - \lambda| \geq 3\lambda) \\ \leq \frac{V[Y]}{9\lambda^2} = \frac{2}{36} = \frac{1}{18} \end{aligned}$$

Example

- Chebyshev can be very pessimistic... (but use very little information!)
- We know that if Z is standard normal then

$$P(|Z| \geq 1.96) \approx 0.05 \qquad P(|Z| \geq 2.58) \approx 0.01$$

or by scaling for $X \sim N(\mu, \sigma^2)$

$$P(|X - \mu| \geq 1.96\sigma) \approx 0.05 \qquad P(|X - \mu| \geq 2.58\sigma) \approx 0.01$$

- By Chebyshev though

$$P(|X - \mu| \geq 1.96\sigma) = P(|Z| \geq 1.96) \leq \frac{1}{1.96^2} = 0.26....$$
$$P(|X - \mu| \geq 2.58\sigma) = P(|Z| \geq 2.58) \leq \frac{1}{2.58^2} = 0.15..$$

Chebyshev and Hoeffding's for binomial RV

We Y a binomial RV with parameters n (number of trials) and p (probability of success). Since $E[Y] = np$ (order n) we are interested in deviations of order $n\epsilon$.

Concentration Inequalities for binomial RV

Suppose Y is a binomial RV with parameters n (number of trials) and p (probability of success).

Chebyshev:

$$P\left(|Y - np| \geq n\epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

Hoeffding:

$$P\left(|Y - np| \geq n\epsilon\right) \leq 2e^{-2n\epsilon^2}$$

The proof of Hoeffding's inequality requires more advanced tools. It is much better than Chebyshev (especially when n is large..).
Note that $p(1-p)$ is less than $\frac{1}{4}$ (since the maximum is at $p = \frac{1}{2}$). Useful when σ is not known.

Example

- You math exam consists of 25 multiple choice exams with 5 possible answers.
- You despise your professor and his stupid french accent and pick the answers at random.
- What are the intervals $[\mu - 2\sigma, \mu + 2\sigma]$ and $[\mu - 3\sigma, \mu + 3\sigma]$
- The exam is curved and you will pass if you score exceed 50%. Estimate the probability you pass the exam.

Application: randomized algorithm

This is a common set-up in machine learning and artificial intelligence: We build an algorithm to perform a certain task. For example image recognition.

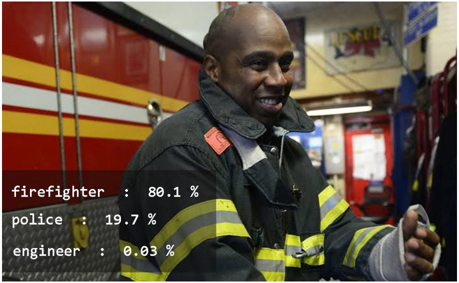

Figure: Image recognition: see this blog post for details and code

To test the performance of the algorithm we use a training set of size n and use human intelligence (we label the picture with the true result!)

We measure the quality of the algorithm by computing

$$\widehat{p}_n = \frac{\text{\# of correctly identified picture}}{n} = .978$$

How much should we trust this result?

We build a probabilistic model and consider a binomial $Y \sim B_{n,p}$ with p unknown. Then

$$\widehat{p}_n = \frac{Y}{n} = \text{ approximate probability from the validation set}$$

$$p = \text{true probability of an image being identified}$$

The quantity $\widehat{p}_n$ is an approximation from the data to the true value p.

Chebyshev

$$P\left(|Y - np| \le n\epsilon\right) = P\left(-n\epsilon \le Y - np \le n\epsilon\right) = P\left(\frac{Y}{n} - \epsilon \le p \le \frac{Y}{n} + \epsilon\right)$$

So by Chebyshev

$$P\left(\widehat{p}_n - \epsilon \le p \le \widehat{p}_n + \epsilon\right) \ge 1 - \frac{\sigma^2}{n\epsilon^2}$$

Set

- $\delta =$ confidence level. ($\delta = 0.01$ means 99% confidence level)
- $\epsilon =$ precision.
- $n =$ sample size, (e.g. $n = 20,000$ (size of the validation set)

$$\frac{1}{4n\epsilon^2} = \delta = 0.01 \iff \epsilon = \sqrt{\frac{1}{4n\delta}} = \sqrt{\frac{100}{4 \times 20,000}} = \sqrt{\frac{1}{800}} = 0.03535...$$

So Chebyshev says that with 99% confidence

$$p \in [.978 - 0.03535, .978 + 0.3535]$$

Hoeffding

By Hoeffding we find that

$$P\left(\frac{Y}{n} - \epsilon \leq p \leq \frac{Y}{n} + \epsilon\right) \geq 2e^{-2n\epsilon^2} = \delta$$

so if we set

$$2e^{-2n\epsilon^2} = \delta \Longleftrightarrow \epsilon = \sqrt{\frac{\ln(2/\delta)}{2n}} = \sqrt{\frac{\ln(200)}{40,000}} = 0.011$$

Better than Chebyshev...