Worksheet 20

Bounding Probabilities

Simple intuition:

1. Draw the normal pdf. Highlight the portion of the pdf capturing $\{|X - \mu| \geq k\sigma\}$ for $k = 0.5, 1, 2, 5$, roughly.

2. If X and Y are two different random variables, is it possible for Chebyshev to yield the exact same bound for them?

Solution: Yes, as it only depends on the first two moments.

3. What are some reasons Chebyshev may be lossy? What are some reasons it may be sharp?

Solution: When tails are extremely light (e.g. Gaussian or Exponential), Chebyshev will be pessimistic. There are many more heuristics, but this is a particularly relevant one for us.

Calculations:

1. Suppose X is now Poisson with parameter λ. What are μ and σ for this distribution?

   (a) Compute $\mathbb{P}[|X - \mu| > 2 \cdot \sigma]$.

   Solution: Simply add up all values outside the box $\lambda - 2\sqrt{\lambda} to + 2\sqrt{\lambda}$. This looks something like

$$\sum_{k=0}^{\lfloor \lambda - 2\sqrt{\lambda} \rfloor} f(k) + \sum_{k=\lceil \lambda + 2\sqrt{\lambda} \rceil}^{\infty} f(k)$$

   The specifics of the formula aren't as important as the basic idea.

   (b) Approximate $\mathbb{P}[|X - \mu| > 2 \cdot \sigma]$ using Chebyshev.

   Solution This is of course very easy – it's less than $\frac{1}{4}$.

   (c) Approximate $\mathbb{P}[|X - \mu| \leq 0.5 \cdot \sigma]$ using Chebsyhev.

   Solution Here the bound is quite silly, we only know it is ≥ -3. This will be true of any $k < 1$.

2. Suppose that X has Laplace distribution with mean 0, i.e. its pdf is

$$f(x) = \frac{1}{2} e^{-|x|}.$$

   Note that the variance of this distribution is 2.

   (a) Compute $\mathbb{P}[|X| > 4]$.

Solution This is now an integral over the region $(-\infty, -4)$ and $(4, \infty)$, so the probability is equal to

$$\begin{aligned}\frac{1}{2}\int_{-\infty}^{-4} e^{-|x|}dx + \frac{1}{2}\int_{4}^{\infty} e^{-|x|}dx &= \frac{1}{2}\int_{-\infty}^{-4} e^{x}dx + \frac{1}{2}\int_{4}^{\infty} e^{-x}dx \\ &= \int_{4}^{\infty} e^{-x}dx \\ &= \frac{1}{e^4}.\end{aligned}$$

Note this is a good chance to practice using absolute values and integrating over such regions!

(b) Compute $\mathbb{P}[|X| \geq 4]$.

Solution Just pointing out it's the same thing for continuous distributions.

(c) Use Chebyshev to approximate $\mathbb{P}[|X| > 4]$.

Solution One can either write $r = 4$ and get $\frac{4}{16} = \frac{1}{4}$ or one can write $k \cdot \sigma$ for $k = 2$ and $\sigma = 2$ to get $\frac{1}{2^2}$. Potato potato (much better written than spoken).

Major takeaway: Chebyshev gets us out of lots of work, at the expense of not being super precise!

Source: Rosen's *Discrete Mathematics and its Applications*.

Math 10B with Professor Stankova
Worksheet, Discussion #20; Tuesday, 4/9/2019
GSI name: Roy Zhao

Chebyshev's Inequality

Concept

1. Chebyshev's inequality allows us to get an idea of probabilities of values lying near the mean even if we don't have a normal distribution. There are two forms:

$$P(|X-\mu| < k\sigma) = P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

$$P(|X-\mu| \geq r) \leq \frac{Var(X)}{r^2}.$$

The **Pareto** distribution is the PDF $f(x) = c/x^p$ for $x \geq 1$ and 0 otherwise. Then this is a PDF and $c = p-1$ if and only if $p > 1$. The mean exists and $\mu = \frac{p-1}{p-2}$ if and only if $p > 2$. Finally the variance exists and $\sigma^2 = \frac{p-1}{p-3} - (\frac{p-1}{p-2})^2$ if and only if $p > 3$.

Example

2. Let $f(x) = \frac{5}{x^6}$ for $x \geq 1$ and 0 otherwise. What bound does Chebyshev's inequality give for the probability $P(X \geq 2.5)$? For what value of a can we say $P(X \geq a) \leq 15\%$?

Solution: We calculate the mean as

$$\mu = \int_{-\infty}^{\infty} xf(x)dx = \int_1^{\infty} \frac{5}{x^5}dx = \frac{5}{4}.$$

The variance is

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = \int_1^{\infty} \frac{5}{x^4}dx - \frac{25}{16} = \frac{5}{3} - \frac{25}{16} = \frac{5}{48}.$$

In order to use Chebyshev's, we need to have a symmetric probability. Namely, we can only use it for something like $P(|X-\mu| \geq r)$. We write $P(X \geq 2.5) = P(X \geq \frac{5}{4} + \frac{5}{4})$. So write

$$P(|X-\mu| \geq \frac{5}{4}) = P(X \geq 2.5 \text{ or } X \leq 0) = P(X \geq 2.5) \leq \frac{\sigma^2}{(5/4)^2} = \frac{5/48}{25/16} = \frac{1}{15}.$$

If we want $P(X \geq a) \leq 15\% = 0.15 = \frac{15}{100}$ using Chebyshev's inequality, then we need

$$\frac{15}{100} = \frac{\sigma^2}{r^2} = \frac{5/48}{r^2}$$

So

$$r = \sqrt{\frac{5/48}{15/100}} = \sqrt{\frac{100}{3 \cdot 48}} = \frac{10}{12} = \frac{5}{6}.$$

Therefore, we need $P(X \geq \mu + 5/6)$ and $a = \frac{5}{4} + \frac{5}{6} = \frac{25}{12}$.

Problems

3. **TRUE** False Chebyshev's inequality can tell us what the probability actually is.

Solution: Like error bounds, Chebyshev's inequality gives us an estimate and most of the time not the actual probability. But there is one case where it does give us the exact probability.

4. True **FALSE** For Chebyshev's inequality, the k must be an integer.

Solution: We can take k to be any positive real number.

5. **TRUE** False The Chebyshev's inequality also tells us $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$.

Solution: This is the complement probability of the first form of the inequality.

6. True **FALSE** Chebyshev's inequality can help us estimate $P(\mu - \sigma \leq X \leq \mu + \sigma)$.

Solution: Using Chebyshev's inequality, we get that this probability is greater than $1 - 1/1 = 0$ which we knew anyway because it is a probability.

7. **TRUE** False We can use Chebyshev's inequality to prove the Law of Large Numbers.

> **Solution:** We write
> $$\lim_{n\to\infty} P(|\bar{X}-\mu|>\epsilon) \le \lim_{n\to\infty} \frac{Var(\bar{X})}{\epsilon^2} = \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0.$$

8. Let $f(x)$ be $(2/3)x$ from $1 \le x \le 2$ and 0 everywhere else. Give a bound using Chebyshev's for $P(10/9 \le X \le 2)$.

> **Solution:** The mean is 14/9 and so this probability is $P(14/9 - 4/9 \le X \le 14/9 + 4/9)$. Letting $4/9 = k\sigma = k\sqrt{26}18$, we calculate that $k = \frac{8}{\sqrt{26}}$. Then, we have that $P(10/9 \le X \le 2) \ge 1 - 1/k^2 = 1 - 1/(64/26) = \frac{19}{32}$.

9. Let $f(x)$ be the uniform distribution on $0 \le x \le 10$ and 0 everywhere else. Give a bound using Chebyshev's for $P(2 \le X \le 8)$. Calculate the actual probability. How do they compare?

> **Solution:** The mean is 5 and so this probability is $P(5-3 \le X \le 5+3)$. Letting $3 = k\sigma = k5\sqrt{3}/3$, we calculate that $k = \frac{9}{5\sqrt{3}}$. Then, we have that $P(2 \le X \le 8) \ge 1 - 1/k^2 = 1 - 1/(81/75) = \frac{2}{27}$.
>
> The actual probability is $\frac{6}{10} = \frac{3}{5}$ which is much more than $\frac{2}{27}$.

10. Let $f(x) = e \cdot e^x$ for $x \le -1$ and 0 otherwise. Give a bound using Chebyshev's for $P(-4 \le X \le 0)$. For what a can we say that $P(X \ge a) \ge 0.99$?

> **Solution:** Since the mean is -2 and the standard deviation is 1, using Chebyshev's inequality, we have that $P(-4 \le X \le 0) = P(\mu - 2\sigma \le X \le \mu + 2\sigma) \ge 1 - \frac{1}{2^2} = \frac{3}{4}$. So an estimate would be $\frac{3}{4}$. The real answer is ≈ 0.95.
>
> In order for $P(X \ge a) \ge 0.99$, we set $a = \mu - k\sigma$ and we know that $P(\mu - k\sigma \le X \le \mu + k\sigma) \ge 1 - \frac{1}{k^2}$. We want to set this lower bound to 0.99 and doing so gives $k = 10$. Thus, we have that $P(\mu - 10\sigma \le X \le \mu + 10\sigma) = P(-12 \le X \le 8) = P(-12 \le X) \ge 0.99$. So, we have that $a = -12$.

11. Let $f(x)$ be $4/x^5$ for $x \ge 1$ and 0 everywhere else. Give a bound using Chebyshev's for $P(X \le 3)$.

Solution: The mean is $4/3$ and since $f(x) = 0$ for all $x < 1$, we have that $P(X \leq 3) = P(4/3 - 5/3 \leq X \leq 4/3 + 5/3) \geq 1 - 1/k^2$. Here, we have that $5/3 = k\sigma = k\sqrt{2}/3$ and so $k = 5/\sqrt{2}$ and $1 - 1/k^2 = 1 - 1/(25/2) = \frac{23}{25}$.