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 $\mathbb{P}[|X| > 4]$.

Worksheet 20

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*.