International Journal of Business Analytics and Intelligence
6 (1) 2018, 01-03
http://publishingindia.com/ijbai/

Analytically Yours:

On Tails

Arnab Kumar Laha*

This article is about tails. Not only animals have tails, but probability distributions have tails too. We will be discussing the tails of probability distributions in this article. In recent times, we have heard about different kinds of tails of probability distributions such as light tail, heavy tail, and long tail. What are these and what should we know about them? This article aims to give a glimpse.

A typical course in probability and statistics begins with definitions of measures of central tendency and variation. One of the most popular measures of central tendency of a given dataset $\{x_1,\ldots, x_n\}$ is the arithmetic mean $(\bar{x})$ which is defined as $\bar{x} = \frac{x_1 + \ldots + x_n}{n}$. When considering a random variable X having a cumulative distribution function (cdf) F_X, we define the analogue of the arithmetic mean called the Expectation of as $E(X) = \int_{-\infty}^{\infty} x dF_X(x)$. For discrete random variables with probability distribution $\{(x_i, p_i): i = 1,\ldots\}$ where $p_i = P(X = x_i)$ and $\sum_{i=1}^{\infty} p_i$ the above expression of $E(X)$ simplifies to give $E(X) = \sum_{i=1}^{\infty} x_i p_i$. Again for absolutely continuous random variables having probability density function (pdf) f_X, we get $E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$. Now, suppose we have information about $E(X)$, can we say something useful about the cumulative distribution function (cdf) F_X (recall that $F_X(t) = P(X \le t)$? Markov's inequality provides an answer to this question for positive random variables.

Suppose, the random variable is positive. Then, the Markov inequality states that for any $t > 0$, $P(X > t) \le \frac{E(X)}{t}$ i.e. $F(t) = P(X \le t) \ge 1 - \frac{E(X)}{t}$. As an example suppose $E(X) = 4$; then, we can say that $P(X > 16) \le 0.25$ and $P(X \le 100) \ge 0.96$. More generally, we can say that $P(X \le c.E(X)) \ge 1 - \frac{1}{c}$ for any $c > 1$.

The most widely used measure of variation of a dataset is standard deviation (sd) which is defined as $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$. It is often convenient to work with s^2 (called the sample variance) instead of s. For a random variable X, its variance is defined analogously as:

$$\mathrm{Var}(X) = E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - E(X))^2 \, dF_X(x)$$

In the discrete case, the above is interpreted $\mathrm{Var}(X) = \sum_{i=1}^{\infty} (x_i - E(X))^2 p_i$; whereas in the absolutely continuous case, this is interpreted as $Var(X)$ $\int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx$. What can we say about F_X if we know both $E(X)$ and $Var(X)$? Can we do better than the case when we only knew $E(X)$?

The Chebyshev's inequality tells us that indeed we can generally do much better if we have information about both the expectation and variance of the random variable X. Noting $P(|X - E(X)| > t) = P((X - E(X))^2 > t^2)$ and then applying the Markov's inequality we get:

$$P(|X - E(X)| > t) \le \frac{E((X - E(X))^2)}{t^2} = \frac{Var(X)}{t^2}$$

Assume as before $E(X) = 4$ and suppose that we have the additional information that $sd(X) = 2$, i.e. $Var(X) = 4$

Now, $P(X > 16) = P(X - 4 > 12) \le P(|X - 4| > 12) \le \frac{2^2}{12^2}$ $= 0.028$ which is a vast improvement on the upper bound of 0.25 we obtained earlier. Taking $t = c.sd(X)$, we get, $P(|X - E(X)| > c.sd(X)) \le \frac{1}{c^2}$

* Indian Institute of Management, Ahmedabad, Gujarat, India. Email: arnab@iima.ac.in

Next we look at Cantelli's inequality which is a generalization of Chebyshev's inequality. Let us denote $E(X) = \mu$ and $sd(X) = \sigma$. Hence, $Y = X - \mu$ has expectation 0 and variance σ^2. Then, for $t, u \geq 0$, we get

$$\begin{aligned} P(X-\mu > t) &= P(Y > t) \\ &= P(Y + u > t + u) \\ &= P((Y+u)^2 > (t+u)^2) \\ &\leq \frac{E((Y+u)^2)}{(t+u)^2} \text{ (by Markov inequality)} \\ &\leq \frac{\sigma^2 + u^2}{(t+u)^2} \end{aligned}$$

Since the above inequality is true for every value of $u \geq 0$, we get,

$$P(X-\mu > t) \leq \inf_{u \geq 0} \frac{\sigma^2 + u^2}{(t+u)^2)}$$

Using calculus, it is easy to show that $\inf_{u\geq 0} \frac{\sigma^2+u^2}{(t+u)^2} = \frac{\sigma^2}{\sigma^2+t^2}$ yielding the Cantelli's inequality:

$$P(X-\mu > t) \leq \frac{\sigma^2}{\sigma^2+t^2}.$$

Now, if we apply Cantelli's inequality to find an upper bound of $P(X > 16)$ where the random variable X has $E(X) = 4$ and $sd(X) = 2$, we get,

$$P(X > 16) = P(X - 4 > 12) \leq \frac{2^2}{2^2 + 12^2} = 0.$$

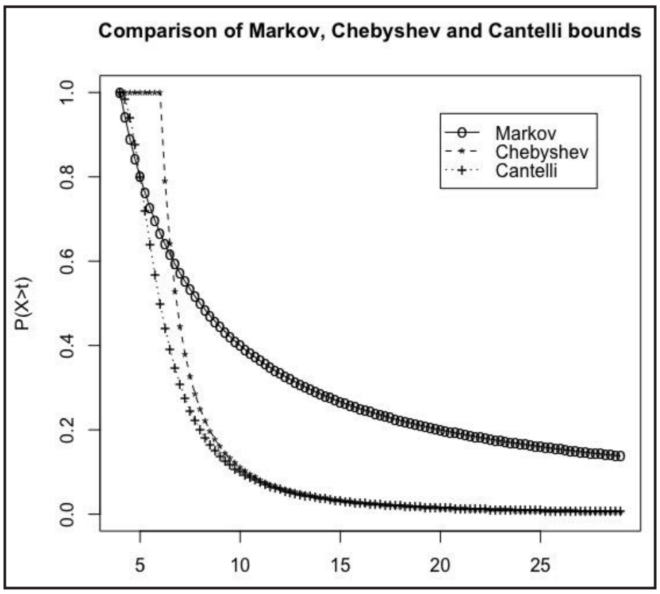

Fig. 1: Comparison of Markov, Chebyshev and Cantelli Bounds

In Figure 1, we give a comparison of the three bounds for $P(X > t)$ discussed above for different values of t.

You may recall that the moment generating function (mgf) of the random variable X is defined as $M_X(\lambda) = E(e^{\lambda X})$. The distributions for which the mgf exists are called *light-tailed*. Normal, Exponential, Uniform, and Gamma are some commonly occurring light-tailed distributions. Since the exponential function is strictly increasing, for $\lambda > 0$ we get using Markov inequality:

$$P(X > t) = P(\lambda X > \lambda t) = P(e^{\lambda X} > e^{\lambda t}) \leq \frac{E(e^{\lambda X})}{e^{\lambda t}} = \frac{M_X(\lambda)}{e^{\lambda t}}.$$

Since this inequality holds for every $\lambda > 0$, we get,

$$\begin{aligned} P(X > t) \leq \inf_{\lambda>0} \frac{M_x(\lambda)}{e^{\lambda t}} &= \inf_{\lambda>0} e^{\ln M_X(\lambda) - \lambda t} \\ &= e^{\inf_{\lambda>0}(\ln M_X(\lambda) - \lambda t)} \end{aligned}.$$

The above inequality is referred to as the Chernoff's inequality.

As an example suppose $X \sim N(0, \sigma)$. Then, $M_X(\lambda) = e^{\frac{1}{2}\lambda^2\sigma^2}$. Now using calculus, it can be easily seen that $e^{\inf_{\lambda>0}(\ln M_X(\lambda)-\lambda t)} = e^{-\frac{t^2}{2\sigma^2}}$ which implies $F(t) = P(X \leq t) \geq 1 - e^{-\frac{t^2}{2\sigma^2}}$

A distribution is said to be *heavy-tailed* if the mgf $M_X(t)$ does not exist (i.e., is not finite for any $t > 0$). Pareto, Cauchy, and Log-normal distributions are some examples of heavy-tailed distributions that occur frequently in Finance as distribution of returns of stocks. Note that a distribution which is not heavy-tailed is light-tailed. Table 1 gives exceedance probabilities for both normal and Cauchy distributions with Median = 0 and MAD = 1 (recall MAD is the acronym for Median Absolute Deviation about median). For normal distribution, the standard deviation is related to MAD through the relation $\sigma = 1.4826 MAD$. In Table 1, we observe that the exceedance probabilities $P(X > x)$ for the Cauchy distribution are much larger than those of the normal distribution. While the chance of observing a departure of magnitude greater than 15 MAD from the median is practically 0 for the normal distribution, it is approximately 2% for the Cauchy distribution. This suggests that a careful study of the tail of a probability distribution is very important in many business applications such as determination of Value-at-Risk of market investments.

Now suppose, $X \sim F$ satisfies the condition that for any fixed $y > 0$, $P(X > x + y | X > x) \to 1$ as $x \to \infty$, i.e., $\frac{1-F(x+y)}{1-F(x)} \to 1$ as $x \to \infty$. Then, the distribution F is said to be long-tailed. It can be proved that every long-tailed distribution is heavy-tailed. However, the converse is not true in general. Some interesting properties of random variables having long-tailed distributions are as follows:

(a) Let $X \sim F$ and $Y \sim G$ be two independent random variables and suppose that the distributions F and G are both long-tailed. Then, the distribution of $X + Y$ is also long-tailed.

(b) Suppose $X_1,\ldots, X_n$ are independent identically distributed random variables having a long-tailed distribution then the distributions of (i) $max\{X_1,\ldots, X_n\}$, (ii) $min\{X_1,\ldots, X_n\}$, and (iii) $X_1 + \ldots + X_n$ are all long-tailed.

Table 1: Exceedance probabilities for Normal and Cauchy distributions with median = 0 and MAD = 1

x	$P(X > x)$ *for Normal*	$P(X > x)$ *for Cauchy*
3	2.1×10^{-2}	0.102
4	3.5×10^{-3}	0.078
5	3.7×10^{-4}	0.063
6	2.6×10^{-5}	0.053
8	3.4×10^{-8}	0.040
10	7.7×10^{-12}	0.032
15	0	0.021
20	0	0.016