Math Notes

Kieran Harvie

Contents

0.1 Variance, and the arbitrariness thereof

For some time I have wondered about the arbitrariness around the use of mean in the definition of variance. For example why the arithmetic mean instead of the geometric or root-mean-squared? And also why do square root the variance to give the standard deviation?

Well the strictness of Chebyshev's inequality might provide a reason.

Conditional Expected Value

This proof of Chebyshev's inequality relies of the conditional expected value, so just to reiterate:

$$E[X|X \geq a] \geq a$$

Since everything X can be is greater then a it's expected value must be greater than a. Notice the strictness of the inequality, this will be used to make the following inequalities much strict.

Chebyshev

Consider the following inequality:

$$\begin{aligned}
&E[(X-a)^2] \\
=&P(|X-a| \leq b)E[(X-a)^2||X-a| \leq b] + P(|X-a| > b)E[(X-a)^2||X-a| > b] \\
\geq&P(|X-a| \leq b) \cdot 0 + P(|X-a| > b)b^2 \\
&\Rightarrow \frac{E[(X-a)^2]}{b^2} \geq P(|X-a| > b)
\end{aligned}$$

We can see that the bound would be better if we pick a such that $E[(x-a)^2]$ is minimized.

$$\begin{aligned}
E[(X-a)^2] =&E[((X-\mu)-(a-\mu))^2] \\
=&E[(X-\mu)^2 - 2(X-\mu)(a-\mu) + (a-\mu)^2] \\
=&E[(X-\mu)^2] - 2(a-\mu)E[X-\mu] + E[(a-\mu)^2] \\
=&\sigma^2 + (a-\mu)^2
\end{aligned}$$

Hence the bound is best at $a = \mu$.

A General Relation

This proof can be generalized, Assume:

$$f(S') \geq 0, \quad g(S) \geq 0$$

Then through:

$$E[f(X)] = P(X \in S)E[f(X)|X \in S] + P(X \in S')E[f(X)|X \in S']$$

We have:

$$1 - \frac{E[g(X)]}{E[g(X)|X \in S']} \leq P[X \in S] \leq \frac{E[f(X)]}{E[f(X)|X \in S]}$$

With dual equality if:

$$f = 1_S, \quad g = 1_{S'}$$

Markov

A similar inequality can be found:

$$\begin{aligned}
\mu =& E[X] \\
=& P(X \leq a)E[X|X \leq a] + P(X \geq a)E[X|X \geq a] \\
\geq& 0 \cdot E[X|X \leq a] + P(X \geq a)a \\
\frac{\mu}{a} \geq& P(X \geq a)
\end{aligned}$$

Covariance

Lets try to find the lest squares regression between X and Y such that:

$$E[X] = E[Y] = 0, E[X^2] = E[Y^2] = 1$$

Since the expected values are both zero the line is through the origin

$$\begin{aligned}
\sum_n (mx_n + c - y_n)^2 =& nE[(mX + c - Y)^2] \\
=& n\left(E[m^2X^2] + E[c^2] + E[Y^2] + E[2cmX] + E[-2mXY] + E[-2cY]\right) \\
=& n(m^2 + c^2 + 1 - 2mE[XY])
\end{aligned}$$

Trying to minimize this value by our selection of trivially gets:

$$c = 0, \quad m = E[XY]$$

Just expanding the definitions gives:

$$COV[X, Y] = E[(X - E[X])(Y - E[Y])] = E[XY] = m$$

Hence the covariance can 'naturally' be interpreted and the first order function between the valuables.

$$E[f(X)] \approx E[f_0 + f_1 X + f_2 X^2/2] = f_0 + \mu f_1 + \sigma^2 f_2/2$$

$$E[f(X)] \approx E[f(\mu) + (X - \mu)f'(\mu) + (X - \mu)^2/2f''(\mu)] = f(\mu) + \frac{f''(\mu)}{2}\sigma^2$$

0.2 Interest Identities

Let P be the principle invested at a rate of r. Consider four different investment scenarios:

- Not invested: $P_0 = P$.
- Fully Invested at the beginning, one instalment at the end:

$$P_1 = (1+r)P$$

- Continuously invested, continuous installments:

$$P_2 = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{P}{n}\left(1+\frac{r}{n}\right)^k = \frac{\exp(r)-1}{r}P$$

- Fully Invested at the beginning, continuous instalments:

$$P_3 = \lim_{n\to\infty} P\left(1+\frac{r}{n}\right)^n = \exp(x)P$$

Interestingly the relative size of P_2 and P_1 depend on r. P_1 starts better than P_2 but switches as r increases.

P_3 is always the best, the proof for P_1 and P_0 are obvious. $P_3 > P_2$ follows from:

$$0 < \int_0^r t\exp(t)\,dt = \big[(t-1)\exp(t)\big]_0^r = (r-1)\exp(r)+1$$

The following interesting identities hold:

$$P_3 = rP_2 + P_0$$

$$P_3 - P = r(P_2 - P) + (P_1 - P)$$

The breaks first neatly breaks P_3 into a nice linear sum. The second does similar for the profit of the investment, total yield minus principle.

0.3 Discontinuities in a Non-decreasing Function

Let f be a non-decreasing function.

Define the jump function J as:

$$J(x) = \inf\{f(t)|t > x\} - \sup\{f(t)|t < x\}$$

This function is well defined since the sets are appropriately bound by $f(x)$. And it is clear that $J(d) \neq 0$ iff d is a discontinuity and that J is non-negative.

Let $U = (x_0, x_1)$. For $d_n \in U$ with $n < m \Rightarrow d_n < d_m$ we have:

$$f(x_1) - f(x_0) \geq \sum_k J(d_k)$$

Proof:

$$\begin{aligned}
&\sum_k J(d_k) \\
=&\inf\{f(t)|t > d_n\} - \sup\{f(t)|t < d_0\} + \sum_k \left[\inf\{f(t)|t > d_{k-1}\} - \sup\{f(t)|t < d_k\}\right] \\
\leq&f(d_n) - f(d_0) + \sum_k \left[f(d_{k-1}) - f(d_k)\right] \\
\leq&f(d_n) - f(d_0) \\
\leq&f(x_n) - f(x_0)
\end{aligned}$$

Let $S_n = \{d \in U|J(d) > \frac{1}{n}(f(x_1) - f(x_0))\}$ From the previous inequality there are at most n elements in S_n. Hence:

$$\{d \in U|J(d) > 0\} = \bigcup_n \left\{d \in U|J(d) > \frac{1}{n}(f(x_1) - f(x_0))\right\}$$

Is countable, hence the number of discontinuities of f on U is countable.

By corollary the discontinuities of a non-decreasing function on $\mathbb{R}$ are countable:

Let f be a non-decreasing function on $\mathbb{R}$. Let $X_n = (n-1, n+1)$, clearly $\mathbb{R} = \bigcup_n X_n$[1]. Assume f has an uncountable number of discontinuities then at least one X_n contains uncountable discontinuities. Otherwise there would be a countable set of countable sets of discontinuities, making them countable. But f being non-decreasing function and having an uncountable number of discontinuities in X_n is a contradiction.

[1]Having them overlap simplify the proof by avoiding literal edge-cases.