Math Notes

Kieran Harvie

Contents

0.1 Mean and Variance, and the arbitrariness thereof

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

Well the strictness of Markov's and Chebyshev's might provide a reason. Both rely 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 stricter.

Markov

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

Chebyshev

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

A General Relation

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:

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

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

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

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.

$$E[f(X)] \approx E[f_0 + f_1X + f_2X^2/2] = f_0 + \mu f_1 + \sigma^2 f_2/2$$

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

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 on 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.

0.4 Integrals and Symmetry

0.4.1 Domain Symmetry

Let U be a subset of $\mathbb{R}^n$ and let $\phi : U \to U$ be a function such that:

$$\begin{aligned} \phi(U) &= U \\ |\det \phi'(\mathbf{u})| &= 1 \end{aligned}$$

Basically ϕ is a linear permutation[2] on U, this is a symmetry in the most direct sense. We obtain the following:

$$\begin{aligned} \int_U f(\mathbf{v})\, d\mathbf{v} &= \int_{\phi(U)} f(\mathbf{v})\, d\mathbf{v} \\ &= \int_U f(\phi(\mathbf{u}))|\det \phi'(\mathbf{u})|\, d\mathbf{u} \\ &= \int_U f(\phi(\mathbf{u}))\, d\mathbf{u} \end{aligned}$$

In particular we get:

$$0 = \int_U \big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big)\, d\mathbf{u}$$

This integral is important since a function can be split into a vanishing and non-vanishing part:

$$\begin{aligned} f(\mathbf{u}) =& \frac{1}{2}\big(f(\mathbf{u}) + f(\phi(\mathbf{u})) + \frac{1}{2}\big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big) \\ \int_U f(\mathbf{u})\, d\mathbf{u} =& \frac{1}{2}\int_U \big(f(\mathbf{u}) + f(\phi(\mathbf{u}))\, d\mathbf{u} + \frac{1}{2}\int_U \big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big)\, d\mathbf{u} \\ =& \frac{1}{2}\int_U \big(f(\mathbf{u}) + f(\phi(\mathbf{u}))\, d\mathbf{u} \end{aligned}$$

[2]Note that ϕ being a permutation requires that if the magnitude of the determinate is constant it must be unity, this can be seen by setting f to a constant.

For example, consider the classic odd function on an integral centered at 0.

$$\phi(x) = -x$$
$$U = [-1, 1]$$

We get the familiar:

$$\int_{-1}^{1} f(x)\,dx = \frac{1}{2}\int_{-1}^{1} \big(f(x) + f(-x)\big)\,dx$$

The utility of this relation can be seen by applying it to the basis of a class of function. Let $V = \langle 1, x, x^2 \rangle$, this is a basis for all parabolas. Notice that x base element vanishes, simplify the evaluation of integrals.

For a 2-D example recall that for two dimensional change of variables:

$$(x, y) = \phi(u, v)$$

We have:

$$|\det \phi'(\mathbf{v})| = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

The rotation symmetry for a regular triangle is:

$$(x, y) = \frac{1}{2}(-u - \sqrt{3}v, \sqrt{3}u - v)$$

Obviously the symmetries act like a group and with functions being a vector space we can use group representations.

0.4.2 Function Symmetry

Let f and ϕ be functions such that:

$$f(t) = \phi'(t) f(\phi(t))$$

Then for arbitrary x_0 and x_1 we have:

$$\begin{aligned}
\int_{x_0}^{x_1} f(t)\,dt &= \int_{\phi(x_0)}^{\phi(x_1)} \phi'(t) f(\phi(t))\,dt \\
&= \int_{\phi(x_0)}^{\phi(x_1)} f(t)\,dt \\
&= \int_{x_1}^{\phi(x_1)} f(t)\,dt + \int_{\phi(x_0)}^{x_1} f(t)\,dt \\
\int_{x_0}^{\phi(x_0)} f(t)\,dt &= \int_{x_1}^{\phi(x_1)} f(t)\,dt
\end{aligned}$$

Hence the integral value is independent of x_n, in particular if ϕ has a fixed point then the integral is zero.

0.5 Lagrange Multiplier

Recall that local extrema x of the function $f : \mathbb{R}^n \to \mathbb{R}$ subject to contrasts g_i satisfy:

$$\nabla f(x) = \sum_i \lambda_i \nabla g_i(x)$$

The core observation is that if ∇f has a component outside the span of ∇g_i then you can move in that direction while keeping g_i's constant, contradicting the point being an extrema.

But the actual constants λ_i have a useful interpretation as the rate the value of f at the extrema changes as the constant g_i changes. To see this pick a particular g_j and construct a d such that:

$$d \cdot \nabla g_i = D\delta_{i,j}$$

You can achieve this by iteratively removing components in some matter like the following:

$$d_0 = \nabla g_0,\ d_{n+1} = d_n - d_n \cdot \nabla g_n$$

Now scale d down such that functions around the extrema can be approximated through targets[3]. We have:

$$\begin{aligned}
g_i(x+d) =& g_i(x) + d \cdot \nabla g_i(x) \\
=& g_i(x) + D\delta_{i,j} \\
f(x+d) =& f(x) + d \cdot \nabla f(x) \\
=& f(x) + \sum_i \lambda_i d \cdot \nabla g_i(x) \\
=& f(x) + D\lambda_j \\
\nabla f(x+d) =& \nabla(f(x) + D\lambda) \\
=& \nabla f(x) \\
=& \sum_i \lambda_i \nabla g_i(x) \\
=& \sum_i \lambda_i \nabla \big(g_i(x+d) - D\delta_{i,j}\big) \\
=& \sum_i \lambda_i \nabla g_i(x+d)
\end{aligned}$$

[3] Those so inclined are free to chase $\epsilon - \delta$'s

From the these equation we can see that $x + d$ satisfy the requirement to be an extrema. We can also see that a change of D in g_i created a change of $\lambda_j D$ in the value at the extrema, hence giving a rate of change of λ_j.

To-Do: Add and example of minimizing height when the two contrasts are parabolic and linear. (You will need to use logs to get the change for the parabola to be a change in it's width and not height.)