SHAPE-MEMORYLESS PROBABILITY DISTRIBUTIONS

HENRY BOTTOMLEY

ABSTRACT. The exponential and geometric probability distributions have the memoryless property, in that past events do not affect the future distribution. This property can be extended to wider families of distributions if the past is allowed to change the scale but not the shape of the future distribution.

1. INTRODUCTION

The property of memorylessness in probability distributions involves conditional probability distributions about the future given that an event has not occurred in the past being the same as the original prior probability distribution. For example, if the lifetime of an unstable atomic particle is seen as a random variable with an exponential probability distribution with a known parameter, and it fails to decay in a certain time, then the probability of its future lifetime follows the same distribution; it has no memory of its failure to decay in the past. A non-negative random variable X has a memorylessness probability distribution provided that $\Pr(X > x + y | X > y) = \Pr(X > x)$ or, using survival functions where $S(x) = \Pr(X > x)$, provided that $S(x + y)/S(y) = S(x)$. For an exponential distribution, we have $S(x) = \exp(-kx)$ for some positive k, and so we demonstrate memorylessness using $\exp(-k(x + y))/\exp(-ky) = \exp(-kx)$.

As an illustration of memorylessness, consider the probability that a random variable with an exponential distribution exceeds its expected value. In effect we have to calculate

$$(1.1) \qquad S(E[X]) = \exp\left(-k \int_0^\infty x \, \exp(-kx)\mathrm{d}x\right) = \exp(-1) \approx 0.37$$

but looking at the position beyond a point $y \geq 0$ we get the same result for the conditional remaining expectation and probability

$$(1.2) \quad S(E[X - y | X > y] | X > y) = \frac{\exp\left(-k \int_y^\infty x \, \exp(-kx)\mathrm{d}x - y\right)}{\exp(-ky)} = \exp(-1)$$

so this calculation produces a constant. Are there other distributions which also produce a constant, perhaps with a different value?

One obvious case is a random variable U with a continuous uniform distribution on the interval $[0, c]$. Clearly, by the symmetry of the distribution, the probability it exceeds its expected value is $\frac{1}{2}$. But this remains the case looking beyond any point y with $0 \leq y < c$. So we have found a memoryless property for a distribution which is not conventionally considered to be memoryless.

In one sense we have cheated. The expectation of U is $\frac{c}{2}$, while the conditional expectation given $U > y$ is $\frac{c+y}{2}$ and so the remaining expectation is $\frac{c-y}{2}$, rather less than we originally had. But in another sense we are not cheating, as the remaining distribution beyond y is still uniform; all that has happened is that the scale has changed, while the shape of the remaining distribution has stayed the same. The aim of this note is to find other examples where the shape remains the same, even if the scale changes to maintain that shape.

More generally, since the conditional shape of the conditional distribution is maintained, so too is the conditional value of any scale-independent statistic. In other words, this means that if a positive random variable X has a shape-memoryless distribution and $f[X]$ is a statistic of the distribution where $f[aX] = f[X]$ for all positive a, then $f[X - y|X > y] = f[X]$ for any y in the support of X. This could be anything from the skewness or kurtosis, if they exist, to the probability of being in excess of $1\frac{1}{2}$ times the interquartile range below the first quartile or above the third quartile (often shown outliers in a box-and-whisker plot).

As an illustration of shape-memorylessness on a discrete distribution, this time without a finite expectation, consider the payoffs of the game in the Bernoullis' *St Petersburg game* of $1, 2, 4, 8, 16, \ldots$ with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \ldots$ respectively. This is not quite shape-memoryless, but it would be if the payoffs were reduced by $\frac{1}{2}$ each to $\frac{1}{2}, 1\frac{1}{2}, 3\frac{1}{2}, 7\frac{1}{2}, 15\frac{1}{2}, \ldots$: for example, initially the payoff $1\frac{1}{2}$ with probability $\frac{1}{4}$ is three times as high as the payoff $\frac{1}{2}$ with probability $\frac{1}{2}$ and the payoff $3\frac{1}{2}$ with probability $\frac{1}{8}$ is seven times as high, while if we know the payoff will be over say $1\frac{1}{2}$ then the additional payoff $7\frac{1}{2} - 1\frac{1}{2} = 6$ with conditional probability $\frac{1/16}{1/4} = \frac{1}{4}$ is again three times as high as the additional payoff $3\frac{1}{2} - 1\frac{1}{2} = 2$ with conditional probability $\frac{1}{2}$ and the additional payoff $15\frac{1}{2} - 1\frac{1}{2} = 14$ with conditional probability $\frac{1}{8}$ is again seven times as high. This pattern would continue looking at other payoffs and different cutoff points.

2. SHAPE-MEMORYLESSNESS FOR CONTINUOUS DISTRIBUTIONS

We start with a standard definition of memorylessness.

Definition 1. *A positive random variable X has a memoryless probability distribution if for $x, y \geq 0$*

$$(2.1) \qquad \Pr(X > x + y | X > y) = \Pr(X > x)$$

or equivalently in terms of survival functions where $S(x)$ means $\Pr(X > x)$

$$(2.2) \qquad \frac{S(x + y)}{S(y)} = S(x)$$

for all x, y in the support of X.

Consider graphically what this means. $S(x + y)$ is what remains of the survival function beyond y. As a function of x it is lower than $S(x)$ because survival functions decrease. But multiplying it by $1/S(y)$ makes it the same as $S(x)$.

For a shape-memoryless distribution we are allowed slightly more flexibility. As a function of x, $S(x + y)/S(y)$ need not be equal to $S(x)$ but it must have the same shape. Our only freedom of manoeuvre is a change of scale in x, stretching or compressing the survival function say by a factor r which may depend on y. So we get the following definition.

Definition 2. *A positive random variable X has a shape-memoryless probability distribution if for $x, y \geq 0$*

$$(2.3) \qquad \frac{S(r(y)x + y)}{S(y)} = S(x)$$

or equivalently in terms of probability

$$(2.4) \qquad \Pr(X > r(y)x + y | X > y) = \Pr(X > x)$$

for all x, y in the support of X and for some scaling function $r(y)$.

The two examples of shape-memoryless probability distributions which we have already considered satisfy that definition: the exponential distribution with survival function $S(x) = \exp(-kx)$ has $r(y) = 1$ while the uniform distribution on $[0, c]$ with $S(x) = 1 - x/c$ has $r(y) = 1 - y/c$. This, combined with taking powers, is enough to give us a family or two of shape-memoryless probability distributions.

Theorem 1. *A shape-memoryless continuous probability distribution of a positive random variable has a survival function either of the form*

$$(2.5) \qquad S(x) = \exp(-kx) \text{ with } x \geq 0 \text{ and } k > 0,$$

or of the form

$$(2.6) \qquad S(x) = (1 - x/c)^{\beta}$$

either with $0 \leq x \leq c$, $\beta > 0$ and $c > 0$, or with $x \geq 0$, $\beta < 0$ and $c < 0$.

Proof. From the definition of shape-memorylessness, we have the functional equation

$$(2.7) \qquad S(x)S(y) = S(r(y)x + y)$$

but we also have $S(y)S(x) = S(r(x)y + x)$ and so $r(x)y + x = r(y)x + y$ requiring either $r(x) = r(y) = 1$ or $x/(1 - r(x)) = y/(1 - r(y))$.

If $r(x) = 1$ then the functional equation reduces to $S(x+y) = S(y)S(x)$ which is the exponential form of Cauchy's functional equation,[1] and given the monotonic decreasing nature of a survival function, this would imply $S(x) = \exp(-kx)$ for some positive k.

Otherwise we have $x/(1 - r(x))$ constant, say equal to non-zero c, thus implying

$$(2.8) \qquad r(x) = 1 - x/c \text{ and } S(x)S(y) = S(x + y - xy/c).$$

Repeating the operation by taking the positive integer n^{th} power of $S(x)$ and setting this equal to $S(z)$ for some z, we get by induction

$$(2.9) \qquad S(z) = S(x)^n = S(c - c(1 - x/c)^n)$$

and thus both

$$(2.10) \qquad z = c - c(1 - x/c)^n \text{ and } x = c - c(1 - z/c)^{1/n}$$

which implies for any positive rational q

$$(2.11) \qquad S(x)^q = S(c - c(1 - x/c)^q)$$

and, since survival functions decrease monotonically, this result can be extended to the positive reals, meaning that knowing c and $S(x_0)$ for any single value of x_0, we can calculate $S(x)$ as

$$(2.12) \qquad S(x) = S(x_0)^{\log(1 - x/c) / \log(1 - x_0/c)}.$$

The functional equation is satisfied mathematically by any

$$(2.13) \qquad S(x) = (1 - x/c)^{\beta}$$

so all we need to do is ensure that such an $S(x)$ is indeed a survival function; for that we need $S(x)$ to be between 0 and 1 and to be decreasing. But its derivative (the negative of the probability density) is

$$(2.14) \qquad \frac{\mathrm{d}S}{\mathrm{d}x} = -(\beta/c)(1 - x/c)^{\beta-1}$$

and so we require β and c to have the same sign, and if those signs are positive we require $0 \leq x \leq c$. Given c and $S(x_0)$ for any single value of x_0 we can calculate $\beta = \log(S(x_0))/\log(1 - x_0/c)$. So, together with the exponential distribution, these provide a complete list of survival functions for continuous random variables meeting the functional equation. $\qquad \square$

If β and c are both positive then we have a probability distribution with finite support on $(0, c]$. This can be seen as a kind of stretched beta distribution where the first parameter is 1. It has mean $c/(1 + \beta)$ and a probability of exceeding the mean of $(1 - \frac{1}{1+\beta})^{\beta}$, which tends towards $\exp(-1)$ from above as β increases.

In contrast, if β and c are both negative as expressed in the theorem then we have a probability distribution with semi-infinite support on $(0, \infty)$; providing $\beta < -1$ we still have mean $c/(1+\beta)$ and the same expression for the probability of exceeding the mean, which tends towards $\exp(-1)$ from below as β grows in absolute magnitude. This distribution could be seen as a kind of stretched beta prime distribution, especially if we were to reverse both signs and write the survival function as $S(x) = (1 + x/c)^{-\beta}$ with β and c both positive.

If we fix the expected value to be μ then the satisfactory survival functions can be written as the exponential $S(x) = \exp(-x/\mu)$ or as

$$(2.15) \qquad S(x) = \left(1 - \frac{x}{\mu(1 + \beta)}\right)^{\beta}$$

with the requirement that $\beta \in (-\infty, -1) \cup (0, \infty)$. As β increases in magnitude towards $\pm\infty$ this converges in distribution to the exponential survival function. The probability of exceeding the mean is

$$(2.16) \qquad S(\mu) = \left(1 - \frac{1}{1 + \beta}\right)^{\beta}$$

which takes every value in $(0, 1)$ as β varies, other than the $\exp(-1)$ provided by the exponential distribution.

For $\beta = 1$, the uniform probability density case, we have survival functions which are linear, while for $\beta = 2$ and $\beta = \frac{1}{2}$ we have survival functions which are arcs of a parabola. For $\beta = -1$ we have survival functions which are part of a hyperbola, and so it should not be a surprise when those particular distributions do not have a mean. If $c > 0$ then c is the maximum value, i.e the smallest value of x where $S(x) = 0$. If β tended to 0 from above, the distribution would tend towards the degenerate case of a single point at $x = c$.

Since we know the derivative of the survival function, we can calculate the hazard function. For an exponential distribution $S(x) = \exp(-kt)$ it is $\lambda(x) = k$. For the other shape-memoryless distributions it is

$$(2.17) \qquad \lambda(x) = \beta/(c - x)$$

FIGURE 1. Illustrations of the survival, density and hazard functions of various shape-memoryless distributions with the same mean

which is part of a hyperbola with an asymptote (rather notionally if c and β are negative) at $x = c$. So hazard functions of this form, or by reversing the signs of the form $\lambda(x) = \beta/(c + x)$, have the shape-memoryless property; the case where $c = 0$ is not a hazard function, unless it it regarded as producing a distribution concentrated at $x = 0$.

If a shape-memoryless distribution has a mean $\mu = c/(1 + \beta)$ then we have $\lambda(x) = \beta/(\mu(1 + \beta) - x)$ and curiously this gives $\lambda(\mu) = 1/\mu$, no matter what the shape of the distribution; this is also the case for exponential distributions since $\mu = 1/k$.

3. Shape-memorylessness for discrete distributions

Ignoring the degenerate case of all probability concentrated at a single point, any shape-memoryless discrete probability distribution must have support on an infinite number of discrete points, as otherwise the conditional distribution would

have support on fewer points than the original distribution and thus would enable them to be distinguished.

A geometric distribution specified by $\Pr(X = n) = (1 - p)p^{n-1}$ for positive integer n and $0 < p < 1$ is memoryless and so shape-memoryless; multiplying all the possible values of X by a constant $a > 0$ would also be a geometric distribution and share the properties.

In general to achieve shape-memorylessness, the probability of each point must be capable of being scaled to earlier points, and so the probabilities of successive points must be in a geometric progression. The gaps between successive points must also be capable of being scaled and so those gaps too must be in a geometric progression. We can state this more precisely in the following theorem.

Theorem 2. *A shape-memoryless discrete probability distribution of a positive random variable has a probability mass function either of the form*

$$(3.1) \qquad \Pr(X = an) = (1 - p)p^{n-1}$$

or of the form

$$(3.2) \qquad \Pr\left(X = a\frac{1 - b^n}{1 - b}\right) = (1 - p)p^{n-1}$$

for some $a > 0$ and $0 < p < 1$, and in the second case $b > 0$ but not 1.

By labeling the successive points of the discrete distribution $x_1, x_2, x_3, \ldots$, this then leads to a survival function either of the form

$$(3.3) \qquad S(x_n) = \exp(-kx_n)$$

for some $k > 0$, or of the form

$$(3.4) \qquad S(x_n) = (1 - x_n/c)^{\beta}$$

for some c and β.

Proof. From the shape-memoryless definition and considering a shift m points up the distribution, we have

$$(3.5) \qquad S(x_{n+m})/S(x_m) = S(x_n) \text{ and } x_{n+m} = r(x_m)x_n + x_m$$

for some scaling function $r(x_m)$. With $m = 1$, $S(x_{n+1}) = S(x_1)S(x_n)$ and letting $p = 1 - \Pr(X = x_1) = S(x_1)$, this implies recursively that $S(x_n) = p^n$ and thus $\Pr(X = x_n) = (1-p)p^{n-1}$. Letting $a = x_1$ and $b = r(x_1)$, and again with $m = 1$, we get $x_{n+1} = bx_n + a$ and so recursively $x_n = an$ when $b = 1$ and $x_n = a(1-b^n)/(1-b)$ otherwise. This gives the first part of the theorem.

If $b = 1$, we can set $k = -\log_e(p)/a$ and get $S(x_n) = \exp(-kx_n)$, while if not then we can set $c = a/(1 - b)$ and $\beta = \log(p)/\log(b)$ so $1 - x_n/c = b^n = p^{n/\beta}$ and we get the familiar looking $S(x_n) = (1 - x_n/c)^{\beta}$. $\square$

Since we have essentially the same survival functions for discrete distributions within their restricted supports, these survival functions continue to satisfy the functional equation. If they have means, then they will be $a/(1-bp)$, slightly greater than the means of the corresponding continuous distributions with essentially the same survival functions. To ensure a meaningful mean we must have $bp < 1$, or equivalently either $b = 1$ or $\beta > 0$ or $\beta < -1$.

FIGURE 2. Illustrations of the left-hand part of the probability mass and survival functions of the variant of the St Petersburg game, together with the survival function of the corresponding continuous distribution

For the variant of the *St Petersburg game* described in the introduction, we have the parameters $a = \frac{1}{2}, b = 2, p = \frac{1}{2}, c = -\frac{1}{2}$ and $\beta = -1$, since

$$(3.6) \qquad \Pr(X = x_{n+m}|X > x_m) = \frac{1/2^{n+m}}{1/2^m} = \frac{1}{2^n} = \Pr(X = x_n) \text{ and}$$

$$(3.7) \qquad x_{n+m} = \frac{2^{n+m} - 1}{2} = (2^m - 1)\frac{2^n - 1}{2} + \frac{2^m - 1}{2} = 2x_m x_n + x_m.$$

With $bp \geq 1$ or equivalently $-1 \leq \beta \leq 0$, and in this particular case $bp = 1$ and $\beta = -1$, there is no mean.

REFERENCES

[1] Cauchy, A.-L., *Cours d'analyse de l'École royale polytechnique. Première partie: Analyse algébrique*, 1821. Chapitre V, problème II, 100-103.
http://gallica.bnf.fr/ark:/12148/bpt6k90195m.image.f104

5 LEYDON CLOSE
LONDON SE16 5PF
E-mail address: se16@btinternet.com
URL: http://www.btinternet.com/~se16/hgb