WARREN D. MACEVOY

LUCK

PERSONAL NOTES

PUBLISHED BY PERSONAL NOTES

WWW.COLORADOMESA.EDU

First printing, October 2016

Contents

List of Figures

List of Tables

Introduction

The point of these notes is to introduce an idea of "luck" that connects mathematical probability with the everyday notion.

These might easily exist, perhaps as another name, in the literature; I just don't know what it is called.

As a motivating problem, imagine walking along the beach and asking a random person to toss a tennis ball so that it lands in the sand. The probability that it lands at some point would depend on the habits of the thrower and the details of the beach, but we can summarize this as some probability distribution, $p(x) = \rho(x)dx$, where $x \in \mathbb{R}^2$ is a suitable coordinate system for the beach in question. It would almost certainly not be a uniform distribution, and it would almost certainly not be particularly concentrated.

Traditional probability feels uncomfortable here. The chances of the ball landing at a given point is zero, and so miraculous. Yet anyone watching this process would only occasionaly be surprized by the outcome.

For a continuous probablity distribution such as this, the chance of the ball landing in some small area dx near x is $p(x) = \rho(x)dx$. But the ball lands at a point, so dx is zero, so the probablity $p(x) = \rho(x)dx$ is zero.

As common (and mundane, not miraculous) such situations are, the language of statistics seems to have difficulty with the notion. Nor is it limited to continuous cases, just when there are a lot of possible outcomes. Such examples lead to non-zero but very small probabilities.

The real motivation of this came from the space of passwords a person might choose from, which is an effectively infinite discrete space.

To distinguish from the more general notion of luck, note that that there is no extrinsic value on an outcome. To say something is "lucky" often means there is some value (different from the probability) associated with outcomes. However, outcomes that are the most valuable are often the least probable, and outcomes of equal probability ought to be equally lucky. In the most extreme case of all equally probable outcomes (uniform probability), every outcome should have a luck of $\frac{1}{2}$.

These observations lead to the following definition of luck:

> The luck $L(x)$ of an outcome x is the probability of getting any outcome y that is more probable than x, plus one-half the probability of getting any outcome y that is equally probable to x.

From the perspective of discussing luck, it is convenient to have few sets: $\Omega(x)$, the outcomes more likely than x, and $\omega(x)$, the outcomes equally likely to x.

Definition 1. Omega. $\Omega(x)$ is set of outcomes more likely than x:

$$\Omega(x) = \{y \mid p(y) > p(x)\} \,. \tag{1}$$

We define $|\Omega(x)|$ as the probability an outcome is in $\Omega(x)$, $|\Omega(x)| = P(y \in \Omega(x))$. In the discrete case, this is

$$|\Omega(x)| = \sum_{y \in \Omega(x)} p(y), \tag{2}$$

and, in the continuous case,

$$|\Omega(x)| = \int_{\Omega(x)} \rho(y)\, dy \,. \tag{3}$$

Definition 2. omega. $\omega(x)$ is the set of outcomes equally likley to x:

$$\omega(x) = \{y \mid p(y) = p(x)\} \,. \tag{4}$$

For the tyical case of many outcomes with different probabilities, $|\omega(x)|$ is small. For example, $|\omega(x)| = 0$ for any multivariate normal distribution.

Similar to $\Omega(x)$, we define $|\omega(x)|$ as the probability an outcome is in $\omega(x)$, $|\omega(x)| = P(y \in \omega(x))$.

In the discrete case, this is

$$|\omega(x)| = \sum_{y \in \omega(x)} p(y), \tag{5}$$

and, in the continuous case,

$$|\omega(x)| = \int_{\omega(x)} \rho(y)\, dy \,. \tag{6}$$

With these definitions in place, we define luck mathematically as follows:

Definition 3. Luck. The luck of an outcome is the probability getting any more likely outcome, plus half the probability of getting any equally likely outcome:

$$L(x) = |\Omega(x)| + \frac{1}{2}|\omega(x)| \,. \tag{7}$$

Properties of Luck.

- Range of luck. $0 \leq L(x) \leq 1$. This ranges from no luck to perfect luck.
- Lucky outcomes. If $L(x)$ is close to 1, then $p(x)$ is comparatively small, and most outcomes would have a higher probability (you are lucky).
- Unlucky outcomes. If $L(x)$ is close to 0, then $p(x)$ is comparatively large, and most outcomes would have a lower probability (you are unlucky).

- Luck on average. $E(L) = \frac{1}{2}$. On average, luck is always 50:50.

We are interested in cases which have many possible outcomes with low but somewhat different probabilities (like the tennis ball on the beach). If the space is well divided (so $\max|\omega| = \max_x|\omega(x)|$ is small), then there are other interesting properties of luck:

In particular $|\omega(x)| = 0$ for the normal, exponential, beta, and gamma distributions. As a worst-case counterexample, the flattest distribution is the uniform distribution, for which $|\omega(x)| = 1$.

- $E(f(L)) = \int_0^1 f(L)dL - \varepsilon$, where $|\varepsilon| \le \max|f''| \cdot \max|\omega|^2/24$.
- For $p \ge 2$, $E(L^p) = 1/(p+1) - \varepsilon$, $0 \le \varepsilon \le p \cdot (p-1)\max|\omega|^2/24$.
- For $0 \le a \le b \le 1$, $E(L \in [a,b]) = b - a - \varepsilon$, $|\varepsilon| \le \max|\omega|$.

The proofs of these come from the midpoint integration rule and thinking about the general case for figure 1 below.

Example 1. *Coins. Suppose we toss 8 fair coins. The probability of getting exactly x heads out of 8 tosses is given by the binomial distribution*

$$p(x) = \frac{8!}{x!(8-x)!}\left(\frac{1}{2}\right)^8. \tag{8}$$

What is the luck associated with this distribution?

x	$p(x)$	$\Omega(x)$	$\lvert\Omega(x)\rvert$	$\omega(x)$	$\lvert\omega(x)\rvert$	$L(x)$
4	0.2734	{}	0.0000	{4}	0.2734	0.1367
3 or 5	0.2188	{4}	0.2734	{3,5}	0.4375	0.4922
2 or 6	0.1094	{3,4,5}	0.7109	{2,6}	0.2188	0.8203
1 or 7	0.0313	{2,3,4,5,6}	0.9297	{1,7}	0.0625	0.9609
0 or 8	0.0039	{1,2,3,4,5,6,7}	0.9922	{0,8}	0.0078	0.9961

Table 1: This is arranged in increasing luck (which is decreasing probability). Getting exactly $x = 4$ heads is unlucky, requiring only $L = 14\%$ luck, while getting $x = 0$ or $x = 8$ heads is almost 100% luck.

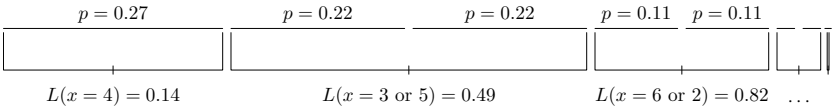

Figure 1: Luck to get x heads on 8 fair coins. After the probabilities are arranged in decreasing order on the unit interval, $L(x)$ is the center point of equally probable outcomes.

Luck on average is $\frac{1}{2}$:

For any distribution, $E(L) = \frac{1}{2}$.

$$E(L) = \sum_{x=0}^{8} p(x) \cdot L(x) = \frac{1}{2}. \tag{9}$$

The second moment should be close to $\frac{1}{3}$:

For any distribution, $E(L^2) = \frac{1}{3} - \varepsilon$, with $0 \le \varepsilon \le \max|\omega|^2/12$.

For the coin distribution, the bound is $0 \le \varepsilon \le 0.016$, and the actual error $\varepsilon = 0.0096$.

$$E(L^2) = \sum_{x=0}^{8} p(x) \cdot L(x)^2 = \frac{1}{3} - 0.0096 \tag{10}$$

The probability luck is in the middle half is about $\frac{1}{2}$:

For any distribution, $E(L \in [a,b]) = b - a + \varepsilon$, with $|\varepsilon| \le \max|\omega|$.

For the coin distribution and $[a,b] = [\frac{1}{4}, \frac{3}{4}]$, $E(L \in [a,b]) = \frac{1}{2} + \varepsilon$ with $|\varepsilon| \le 0.4375$ as the (poor) error bound, and the actual error of $\varepsilon = -0.0625$

$$E(L \in [\frac{1}{4}, \frac{3}{4}]) = \sum_{x=0}^{8} p(x) \cdot \left\{ \begin{array}{ll} 1 & \textit{if } L(x) \in [\frac{1}{4}, \frac{3}{4}] \\ 0 & \textit{otherwise} \end{array} \right\} \\ = \frac{1}{2} - 0.0625\,. \tag{11}$$

Example 2. *Normal. This is a special case of the multivariate normal we cover in the next section, but working out the details for the one-dimensional case can be illuminating. We define the one-dimensional normal distribution with mean μ and variance Σ as*

$$P_{normal}(x; \mu, \Sigma) = \frac{e^{-\frac{(x-\mu)^2}{2\Sigma}}}{\sqrt{2\pi\Sigma}}\,, \tag{12}$$

where

$$\mu = E(x)\,, \tag{13}$$

and

$$\Sigma = E((x-\mu)^2)\,. \tag{14}$$

First note that $\Omega(x)$ is the open the interval between x and x reflected around μ:

$$\Omega(x) = (\min(x, 2\mu - x), \max(x, 2\mu - x))\,, \tag{15}$$

and $\omega(x)$ is the endpoints of that interval:

$$\omega(x) = \{x, 2\mu - x\}\,. \tag{16}$$

Since the 1-dimensional normal distribution is a continuous distribution and $\omega(x)$ is a finite set, $|\omega(x)| = 0$, i.e.,

$$|\omega(x)| = \int_{\omega(x)} P_{normal}(y; \mu, \Sigma)\, dy = 0\,. \tag{17}$$

This means all the luck properties are exact (the error terms are zero), and the $\frac{1}{2}|\omega(x)|$ contributes nothing to the luck of an outcome.

What remains is to calculate luck,

$$L(x) = \int_{\Omega(x)} P_{normal}(y; \mu, \Sigma)\, dy\,. \tag{18}$$

Changing variables to the normalized z-score: $z = \sqrt{\Sigma^{-1}}(x - \mu)$, this can be rewritten as

$$L(x) = \int_{-R}^{R} P_{normal}(y, 0, 1)\, dy\,, \tag{19}$$

where

$$R = |\sqrt{\Sigma^{-1}}(x - \mu)|\,. \tag{20}$$

Using $\mathrm{erf}(x)$*, defined as*

$\mathrm{erf}(x)$ is a normalized integral of the $P_{\mathrm{normal}}(x; \mu = 0, \Sigma = 1/2)$ so that $\mathrm{erf}(0) = 0$ and $\mathrm{erf}(\pm\infty) = \pm 1$.

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy\,, \tag{21}$$

the luck of an outcome can be written as

$$L(x) = \mathrm{erf}\left|\frac{x-\mu}{\sqrt{2\Sigma}}\right|\,. \tag{22}$$

In the next chapter, where we address the more general multivariate normal case, we obtain the approximation,

$$L(x) \approx \frac{1}{2}\left[1 + \mathrm{erf}\left(\left|\frac{x-\mu}{\sqrt{\Sigma}}\right| - \sqrt{\frac{1}{2}}\right)\right]\,. \tag{23}$$

Figure 2 compares the exact and approximate result in the 1-d case.

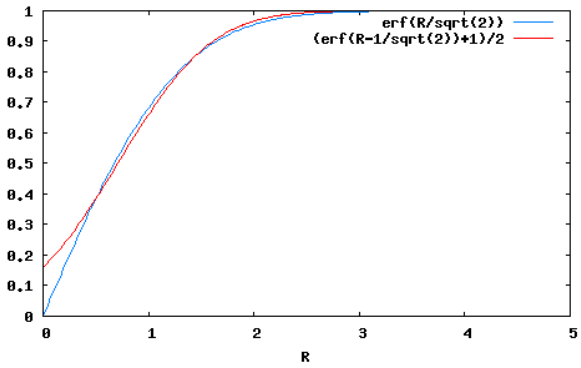

Figure 2: Exact (blue, equation 22) vs approximate (red, equation 23) luck for 1-d normal distribution.

Normal Distribution

Suppose we are in a probability space well approximated by the multivariate normal (Gaussian) distribution of a random variable $x \in \mathbb{R}^n$ with mean μ and non-singular covariance Σ:

The $n = 1$ example in the introduction is just where where μ and Σ are scalars.

$$P_{\text{normal}}(x; \mu, \Sigma) = \frac{e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}}{\sqrt{(2\pi)^n \det \Sigma}}\,, \tag{24}$$

where

$$\mu_i = E(x_i)\,, \tag{25}$$

and

$$\Sigma_{ij} = E((x_i - \mu_i)(x_j - \mu_j))\,. \tag{26}$$

How lucky is some outcome x? From the definition:

$$L(x) = |\Omega(x)| + \frac{1}{2}|\omega(x)|\,, \tag{27}$$

where

$$\Omega(x) = \{y \in \mathbb{R}^n | P_{\text{normal}}(y) > P_{\text{normal}}(x)\} \tag{28}$$

$$= \left\{y \in \mathbb{R}^n \middle| |\sqrt{\Sigma^{-1}}(y-\mu)| < |\sqrt{\Sigma^{-1}}(x-\mu)|\right\} \tag{29}$$

and

$$\omega(x) = \left\{y \in \mathbb{R}^n \middle| |\sqrt{\Sigma^{-1}}(y-\mu)| = |\sqrt{\Sigma^{-1}}(x-\mu)|\right\}\,. \tag{30}$$

Because $\omega(x)$ has no volume in $\mathbb{R}^n$,

$$|\omega(x)| = \int_{\omega(x)} P_{\text{normal}}(y; \mu, \Sigma)\, dy = 0\,. \tag{31}$$

So

$$L(x) = |\Omega(x)| \tag{32}$$

$$= \int_{\Omega(x)} P_{\text{normal}}(y; \mu, \Sigma)\, dy\,. \tag{33}$$

By changing variables to $z = \sqrt{\Sigma^{-1}}(x - \mu)$,

$$L(x) = \int_{|z|<R} P_{\text{normal}}(z; 0, I)\, dz\,, \tag{34}$$

Σ is symmetric and positive definite, and so is its inverse. The square-root can be computed as a Cholesky decomposition. In Scilab, z=**chol**(Sigma)'\(x−mu)

where $R = |\sqrt{\Sigma^{-1}}(x - \mu)|$.

R=**norm**(z)

This can be evaluated in spherical coordinates:

$$L(x) = \frac{1}{\sqrt{(2\pi)^n}} \int_0^R \frac{n\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} r^{n-1} e^{-\frac{1}{2}r^2}\, dr\,, \tag{35}$$

$$= \frac{\gamma(n/2, R^2/2)}{\Gamma(n/2)}\,. \tag{36}$$

The last form uses the lower incomplete gamma function and gamma function, defined to be

L=**cdfgam**("PQ",R^2/2,n/2,1)

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t}\, dt \tag{37}$$

$$\Gamma(s) = \gamma(s, \infty)\,. \tag{38}$$

For any value of n, but particularly for large values, we find the following approximation to be very good:

$$L(x) \approx \frac{1}{2}\left[1 + \text{erf}(|\sqrt{\Sigma^{-1}}(x - \mu)| - \sqrt{n - 1/2})\right]\,. \tag{39}$$

This comes from a Taylor expansion of the log of the integrand in (35), and numerical experimentation on the 1/2 factor. The expansion is specifically invalid for $n = 1$, hence the difference between the general case and the $n = 1$ case (22).

Lapprox=0.5*(1+**erf**(R−**sqrt**(n−1/2)))

Not only is this result pretty, it is very useful. Suppose we have distribution parameters μ and Σ, and would like to know if they fit actual observations. A traditional approach requires a large sample to estimate μ and Σ, but we don't need this, nor do we need to assume that the distribution is normal. We just need to ask if the observations are surprising (lucky or unlucky). In large dimensions, numerical experiments suggest one sample is in most cases sufficient to establish practical certainty (probability of error less than 10^{-15}).

Combining Normal Luck

The approximate result (39) leads to a rule for combining (normal) luck: Suppose there are two independent normal distributions parameterized by $\mu^{(x)}$, $\Sigma^{(x)}$ of dimension n_x, and $\mu^{(y)}$, $\Sigma^{(y)}$ of dimension n_y. What is the luck of a single combined observation (x, y)?

$$L(x, y) \approx \frac{1}{2}\left[1 + \text{erf}(\sqrt{R_x(x)^2 + R_y(y)^2} - \sqrt{n_x + n_y - \frac{1}{2}})\right]\,, \tag{40}$$

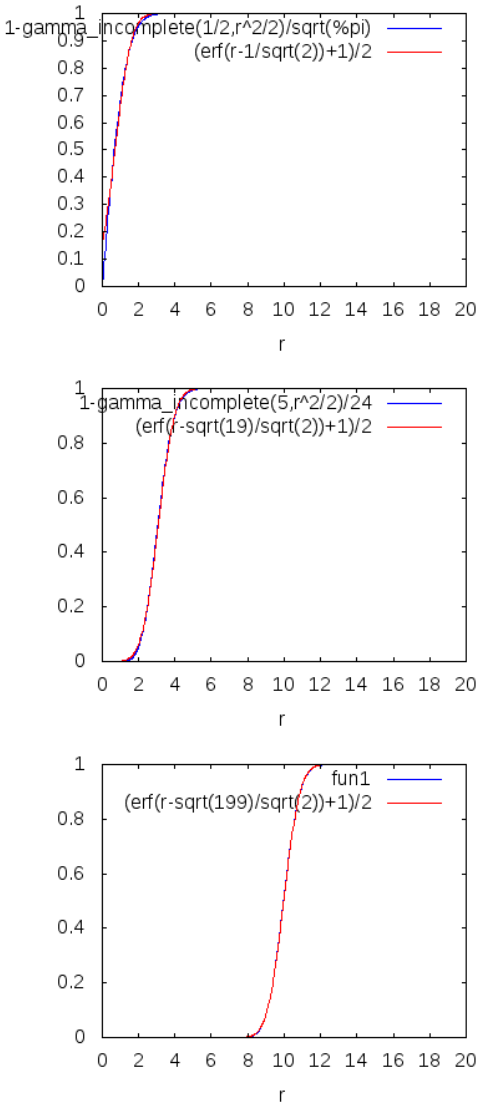

Figure 3: Exact (blue) vs approximate (red) luck for normal distribution for $n = 1$, 10, and 100.

$L^{(x)}(x)$	$L^{(y)}(x)$	$L^{(x)}(y)$	$L^{(y)}(y)$
0.501 417 202 0	1.000 000 000 0	1.000 000 000 0	0.838 180 264 1
0.731 421 266 5	1.000 000 000 0	1.000 000 000 0	0.239 258 143 2
0.982 563 033 9	1.000 000 000 0	1.000 000 000 0	0.271 695 512 7
0.033 455 080 7	1.000 000 000 0	1.000 000 000 0	0.421 320 625 9
0.689 429 934 0	1.000 000 000 0	1.000 000 000 0	0.074 461 655 7
0.736 397 593 7	1.000 000 000 0	1.000 000 000 0	0.294 050 728 4
0.304 521 296 7	1.000 000 000 0	1.000 000 000 0	0.707 849 014 7
0.231 111 574 4	1.000 000 000 0	1.000 000 000 0	0.290 313 093 2
0.585 247 719 9	1.000 000 000 0	1.000 000 000 0	0.636 902 202 8
0.214 552 926 1	1.000 000 000 0	1.000 000 000 0	0.268 989 787 4

Table 2: Luck from two randomly generated distributions $\mu^{(x)}$ and $\mu^{(y)}$ uniformly chosen in $[0,1]^{100}$, and $\Sigma^{(x)}, \Sigma^{(y)}$ are transposed squares of random 100×100 matrices. In each row, x is a sample from the $\mu^{(x)}, \Sigma^{(x)}$, normal distribution, and y is from the $\mu^{(y)}, \Sigma^{(y)}$ distribution. The actual values of x and y are not given, since they are very large (100 numbers each) and uninteresting.

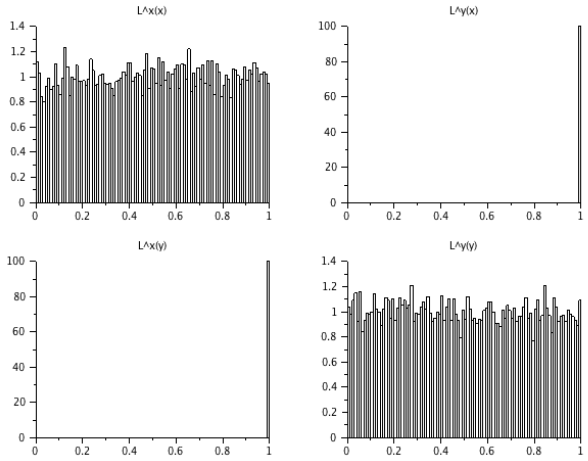

Figure 4: Histograms of 10,000 luck values in the same distributions as table 2. Upshot: y values in the x distribution are viewed as extremely lucky and vice-versa, while they are have uniform luck in their own respective distributions.

where

$$R_x(x) = \text{erf}^{-1}(2L_x - 1) - \sqrt{n_x - \frac{1}{2}} = \sqrt{\left(\Sigma^{(x)}\right)^{-1}}\left(x - \mu^{(x)}\right), \quad (41)$$

$$R_y(y) = \text{erf}^{-1}(2L_y - 1) - \sqrt{n_y - \frac{1}{2}} = \sqrt{\left(\Sigma^{(y)}\right)^{-1}}\left(y - \mu^{(y)}\right). \quad (42)$$

The approximations above, which in the limit are exact, lead to the following natural definition:

Definition 4. Luck-adjusted z-score. For any distribution (not just a normal distribution) with finite mean $\mu = E(x)$ and finite positive definite covarience $\Sigma = E((x-\mu)(x-\mu)^T)$, where x and μ are d_f-dimensional column vectors, it is natural to associate an observation x with the *luck-adjusted z-score*:

$$z_L = \left|\sqrt{\Sigma^{-1}}(x - \mu)\right| - \sqrt{d_f - \frac{1}{2}}. \quad (43)$$

The luck-adjusted z-score from two independent experiments A and B can be combined into one overall score with:

$$\begin{aligned} z_L^{A\times B} = &\sqrt{\left(z_L^A + \sqrt{d_f^A - \frac{1}{2}}\right)^2 + \left(z_L^B + \sqrt{d_f^B - \frac{1}{2}}\right)^2} \\ &- \sqrt{n_A + n_B - \frac{1}{2}}, \end{aligned} \quad (44)$$

and

$$d_f^{A\times B} = d_f^A + d_f^B. \quad (45)$$

Similarly, k repeated independent experiments A can be combined

with

$$z_L^{A^k} = \sqrt{\sum_{i=1}^{k} \left(z_L^{A_i} - \sqrt{d_f^A - \frac{1}{2}} \right)^2} - \sqrt{k \cdot d_f^A - \frac{1}{2}}\,, \tag{46}$$

and

$$d_f^{A^k} = k d_f^A\,. \tag{47}$$

If the combined dimension d_f is large enough that the overal distribution is well approximated by the normal distribution, then the luck associated with the overall set of observations is well approximated by the *normal luck*,

$$L_N = \frac{1}{2}\left[1 + \mathrm{erf}(z_L)\right]\,. \tag{48}$$

If the distribution is in fact normal, the luck can be computed exactly via,

$$L = \frac{\gamma(\frac{n}{2}, \frac{1}{2}(z_L + \sqrt{n - \frac{1}{2}})^2)}{\Gamma(n/2)}\,. \tag{49}$$

There is less than a 0.01 difference between L_N and L for $d_f \geq 22$.

Combining luck-adjusted z-scores is very useful for understanding the implications of multiple experiments. If the model is good, the luck-adjusted z-score will to stay small (within ± 6) as you combine the outcomes of more experiments. If it tends to get large or small, there is something wrong in the model. If z_L tends to negative infinity, this seems to be an indication of a non-stochastic process (it is being gamed), and if z_L tends to positive infinity, the estimates for μ and/or Σ are wrong. The publication of z_L and d_f for experimental results would be very useful for meta analysis, and L_N or L would be a value much easier to interpret for a lay reader.

Scilab Reference Code

The listings in figures 5-7 give Scilab functions for basic luck calculations for multivariate normal distributions.

```
function lnp=mnprobln(x,mu,Sigma)
  [dim,nsamps]=size(x);
  sigma=chol(Sigma)';
  z=sigma\(x-mu*ones(1,nsamps));
  R2=sum(z.^2,'r');
  lnp=-R2/2-(dim/2*log(2*%pi)+...
        sum(log(diag(sigma))));
endfunction
```

Figure 5: Scilab listing to compute the natural log of the probability of multivariate normal outcomes. x is a $n_{\text{dim}} \times n_{\text{samps}}$ of outcomes, μ is a $n_{\text{dim}} \times 1$ column vector of the means, and Σ is a $n_{\text{dim}} \times n_{\text{dim}}$ covariance matrix. The result `lnp` is a $1 \times n_{\text{samps}}$ row vector.

```
function L=mnluck(x,mu,Sigma)
  [dim,nsamps]=size(x);
  one=ones(1,nsamps);
  sigma=chol(Sigma)';
  z=sigma\(x-mu*one);
  R2=sum(z.^2,'r');
  L=cdfgam("PQ",R2/2,dim/2*one,one);
endfunction
```

Figure 6: Scilab listing to efficiently compute the luck of multivariate normal outcomes. x is a $n_{\text{dim}} \times n_{\text{samps}}$ of outcomes, μ is a $n_{\text{dim}} \times 1$ column vector of the means, and Σ is a $n_{\text{dim}} \times n_{\text{dim}}$ covariance matrix. The result L is a $1 \times n_{\text{samps}}$ row vector of the luck associated with each outcome.

```
function L=mnapproxluck(x,mu,Sigma)
  [dim,nsamps]=size(x);
  sigma=chol(Sigma)';
  z=sigma\(x-mu*ones(1,nsamps));
  R=sqrt(sum(z.^2,'r'));
  L=0.5*(1+erf(R-sqrt(dim-0.5)));
endfunction
```

Figure 7: Like `mnluck` in figure 6, but approximates luck via (39).

Multinomial Distribution

Selecting independent samples (with repetition) from a discrete set of outcomes follows the multinomial distribution.

Specifically, select N independent samples (with repetition) from a discrete set of possible outcomes $\{x_i \in \mathbb{Z}_{0+}\}_{i=1}^{n}$ with probabilities p_i, $\sum_{i=1}^{n} p_i = 1$. Not considering order, this will result in $x_i \in Z_{0+}$ counts of outcome w_i, $\sum_{i=1}^{N_p} x_i = N$. The probability of such a sample is given by the multinomial distribution:

$$P_{\text{multinomial}}(x; p) = N! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!} \,. \tag{50}$$

The counts x_i are not independent (they must sum to N), but if the probabilities are small, then the multinomial distribution can be treated as a product of independent Poisson distributions:

$$P_{\text{poisson}}(x; p) = \prod_{i=1}^{N_p} \frac{e^{-p_i} p_i^{x_i}}{x_i!} \,. \tag{51}$$

Approximating the Multinomial as a Normal Distribution

The multinomial distribution can be approximated by a normal distribution,

$$\mu_i = N p_i \,, \tag{52}$$

$$\Sigma_{ij} = \begin{cases} N p_i (1 - p_i) & \text{if } i = j \,, \\ -N p_i p_j & \text{if } i \neq j \,. \end{cases} \tag{53}$$

But the covarience matrix Σ is singular. Because of this, instead of the general normal distribution, the multinomial distribution is approximated by the continuous distribution,

http://www.cnd.mcgill.ca/ ivan/Chelosky-multinomial-decomp-2345957.pdf

$$P(x) = \frac{\exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - N p_i)^2}{N p_i}\right)}{\sqrt{(2\pi)^{(n-1)} n \prod_{i=1}^{n} p_i}} \,. \tag{54}$$

subject to the constraint,

$$\sum_{i=1}^{n} x_i = N \,. \tag{55}$$

Changing coordinates to

$$z_i = \frac{x_i - Np_i}{\sqrt{Np_i}} \tag{56}$$

leads to an integration on the $\mathbb{R}^{n-1}$ hypersurface $n \cdot z = 0$, where

$$n_i = \sqrt{p_i} \tag{57}$$

This has the same form as the multinomial gaussian, except it is one lower dimension, yeilding,

$$L(x) = \frac{\gamma((n-1)/2, |z|^2/2)}{\Gamma((n-1)/2)}\,. \tag{58}$$

Computation

Unfortunately, the explicit computations for luck may be impractical. However, if n independent samples are taken $S = (x_1, \ldots, x_n)$, then it can be estimated as:

$$\begin{aligned}\ell(x) =& \frac{1}{n}\{\#\text{ of outcomes in S more probable than } x\} \\ &+ \frac{1}{2n}\{\#\text{ of outcomes in S equally probable to } x\}\end{aligned}$$

It is a reasonably straightforward calculation to show that

$$E(\ell(x)) = L(x)\,, \tag{59}$$

and

$$E((\ell(x) - L(x))^2) \leq \frac{1}{n}L(x)\cdot(1 - L(x)) \leq \frac{1}{4n}\,. \tag{60}$$

Example 3. *Multinomial. We are aware of no computationally efficient way to exactly compute the luck for a multinomial distribution other than the explicit sum. Suppose there is a questionnaire with 4 possible answers with category probabilities* 0.1, 0.2, 0.3, *and* 0.4. *How lucky would it be to get* 13, 15, 27 *and* 45 *responses of each respective answer in 100 samples?*

The log-probability in this case is the multinomial probability (see figure 8):

$$y = \mathit{mulprobln(x,p)} = \log\left[\left(\sum_{i=1}^{n_p} x_i\right)!\prod_{i=1}^{n_p}\frac{p_i^{x_i}}{x_i!}\right]\,. \tag{61}$$

Numerical estimates of luck requires samples from the given probability distribution. Figure 9 uses a built-in function in scilab to generate such a sample set.

For smaller spaces, luck for a multinomial distribution can be computed explicitly. Figure 10 computes this as an inefficient testing reference.

```
function lnp=mulprobln(x,p)
  [nprobs,nsamps]=size(x);
  pln=log(p);
  y=zeros(1,nsamps);
  for i=1:nsamps
    xi=x(:,i);
    ni=sum(xi);
    lnp(i)=gammaln(ni+1)+...
      sum(xi.*pln-gammaln(xi+1));
  end
endfunction
```

Figure 8: Scilab function to compute the natural log of the probability of multinomial outcomes. p is a $n_{\text{probs}} \times 1$ column vector of category probabilities, x is a $n_{\text{probs}} \times n_{\text{samps}}$ matrix of outcomes, and the result `lnp` is a $1 \times n_{\text{samps}}$ row vector.

```
function x=mulsamp(nsamps,ntrials,p)
  x=grand(nsamps,"mul",ntrials,p(1:(length(p)-1)));
endfunction
```

Figure 9: Scilab listing to get a sample of outcomes from a multinomial distribution. n_{samps} is the number of desired samples, n_{trials} is the number of trials in each sample, and p is a $n_{\text{probs}} \times 1$ column vector of category probabilities. The result x is a $n_{\text{probs}} \times n_{\text{samps}}$ matrix of sample outcomes, where $\sum_{j=1}^{n_{\text{probs}}} x(j,i) = n_{\text{trials}}$.

```
// Used by mulluck below to
// recursively compute luck
function el=mulluckrec(nb,mb,...
                lpx,k,p,y,eps)
  [n,m]=size(lpx);
  el=zeros(n,m);
  for yk=0:nb-mb
     y(k)=yk;
     if k < length(p)-1 then
       el=el+mulluckrec(nb,mb+yk,...
         lpx,k+1,p,y,eps);
     else
       y(length(p))=nb-mb-yk;
       lpy=mulprobln(y,p);
       c=0.5*bool2s(lpy > lpx-eps)+...
            0.5*bool2s(lpy > lpx+eps);
       el = el + c .* exp(lpy);
     end
  end
endfunction

function L=mulluck(x,p,eps)
  if ~exists("eps","local") then
    eps=sqrt(%eps);
  end
  [nprobs,nsamps]=size(x);
  ntrials=sum(x,'r');
  no=min(ntrials);
  n1=max(ntrials);
  assert_checkequal(no,n1);
  L=mulluckrec(no,0,mulprobln(x,p),...
      1,p,zeros(nprobs,1),eps);
endfunction
```

Figure 10: Recursively compute luck of multinomial exactly using exhaustive sum. x is a $n_{\text{probs}} \times n_{\text{samps}}$ matrix of outcomes, and p is a $n_{\text{probs}} \times 1$ column vector of category probabilities. The result is a $1 \times n_{\text{samps}}$ of luck values.

```
function setup=numlucksetup(problns,eps)
  if ~exists("eps","local") then
    eps=sqrt(%eps);
  end

  n=length(problns);
  problns=gsort(problns);

  for pass=1:2
    if pass == 2 then
      setup=zeros(2,count);
    end
    i=1;
    count=0;
    while i<=n
      j=i;
      while j <= n-1 & ...
          problns(i)-problns(j+1) < eps
        j=j+1;
      end
      count=count+1;
      if pass == 2 then
        setup(1,count)= ...
          -0.5*(problns(i)+problns(j));
        setup(2,count)= ...
          (i-1)/n+0.5*(j-i+1)/n;
      end
      i=j+1;
    end
  end
endfunction
```

Figure 11: Given the log of the probabilities of a set of sample data, return a table used for quickly estimating luck. `problns` is a $1 \times n_{\text{nsamps}}$ row vector of logs of probabilities, and `eps` is an optional parameter giving the absolute error (in log space) for considering two probabilities to be equal. Returns `setup`, a $2 \times N$ matrix giving $-\log p(x)$ and estimates for $L(x)$ for each unique probability in the sample. This function is useful generally (not just for multinomial distributions).

```
function L=numluck(problns,setup)
  L=max(0,min(1,interpln(setup,-problns)));
endfunction
```

Figure 12: Estimate luck given log of probabilities and setup. `problns` is a row vector of log-probabilities, `setup` is the setup from a (possibly different) sample.

```
// get sample set
nsamps=10000;
ntrials=100;
p=[0.1;0.2;0.3;0.4];
x=mulsamp(nsamps,ntrials,p);
problns=mulprobln(x,p);

// setup for luck estimates
setup=numlucksetup(problns);

// estimate luck numerically
x0=[13; 15; 27; 45];
problns0=mulprobln(x0,p);
nluck=numluck(problns0,setup);
nsd=sqrt(nluck .* (1-nluck) ./ nsamps);

// optional exact luck
luck=mulluck(x0,p);
sd=sqrt(luck .* (1-luck) ./ nsamps);

// z is approximately normally disributed
z=(nluck-luck) ./ sd;
```

Figure 13: Use the above functions for estimating luck numerically (`nluck`) and estimating the error in luck (numerical standard deviation). The last lines optionally compute the exact values of luck (`luck`) and standard deviations to compare with. One run produced `nluck=0.63025` and `luck=0.62875` with error bound estimates `nsd=0.004827`, `sd=0.0048314` and `z=0.3105`.

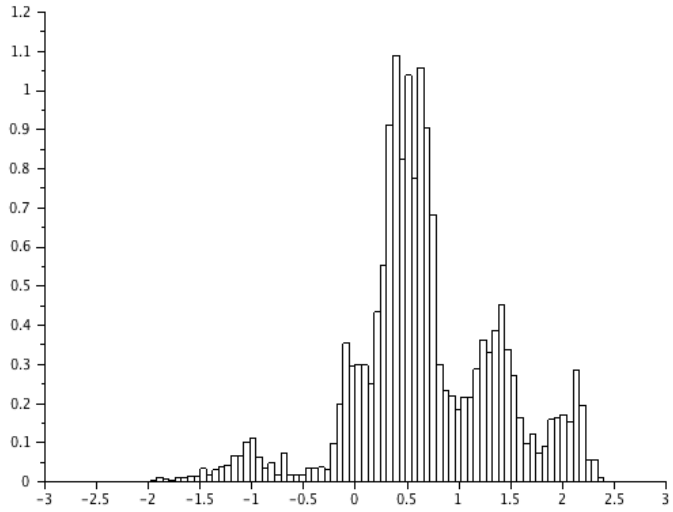

Figure 14: Histogram of z error values for 10,000 numerical approximations of luck. The maximum value of `sd` was `max(sd)=0.005`.

Application - Testing Randomness

A popular randomness testing suite is the Dieharder suite, which can be used to validate various random number generators, including custom generators, against a set of established tests of randomness. Under standard usage, the outcome of each test is a p-value, which if is outside a cutoff $|p - 1/2| > 1/2 - \varepsilon/2$, is considered weak ($\varepsilon = 0.005$) or fails ($\varepsilon = 0.000001$).

The issue is that, assuming the test and source of randomness are correct, the p-value is a uniformly distributed value. This means that ε of the tests will be labeled weak or failed. With many tests and random number generators, there should be failures. This possibility of false negatives makes the results difficult to interpret. On the other hand, a consistently weak random number generator may produce a poor result consistently, just not weak enough to cross the predetermined weak/fail test, so false positives are also a problem.

We took two luck-based approaches to this. The first was to create our own test which we could assess using a luck approach, and the second was to take a luck perspective on the built-in dieharder results.

Max64 Statistics

The first was to design our own test (Max64) with a known exactly computable statistical distribution. The outcomes of each test is combined using the luck-adjusted z-score (43) of individual tests to obtain an overall normal luck score via the repeated use of (44). This is continued until the number of tests exceeds some limit, or z_L reaches a value that constitutes a statistical proof of failure, i.e., $|z_L| > 10$ which should occur less than once in 10^{44} trials.

The chosen statistical test is very simple: the distribution of maximum values on a filtered permutation of the bit stream. This was done because the statistical moments can be computed to high precision (this is crucial, since incorrect moments would result in model failure because of the test, not the source of randomness). The resulting project is hosted as github.com/wmacevoy/testrng, and is

remarkably efficient at determining poor randomness compared to the standard Dieharder suite.

Probability of Maximums

If taking S independent samples $(x_1, \ldots, x_S)$ uniformly of the numbers $\{1 \ldots N\}$, the probability that the maximum value is M is given by

$$P(M) = \text{Prob}(\max_{k=1}^{S} x_k = M) = \left(\frac{M}{N}\right)^S \cdot \left[1 - \left(\frac{M-1}{M}\right)^S\right] \tag{62}$$

$$= \frac{1}{N^S}\left(M^S - (M-1)^S\right). \tag{63}$$

The expected maximum value is given by

$$E(M) = \sum_{M=1}^{N} P(M)M, \tag{64}$$

$$= \frac{1}{N^S}\sum_{M=1}^{N}\left[M^{S+1} - (M-1)^S \cdot ((M-1)+1)\right], \tag{65}$$

$$= \frac{1}{N^S}\left[N^{S+1} - \sum_{M=1}^{N}(M-1)^S\right], \tag{66}$$

$$= N - \sum_{M=1}^{N-1}\left(\frac{M}{N}\right)^S, \tag{67}$$

$$= N - \frac{1}{N^S}H(N-1, -S). \tag{68}$$

Here, we have used the generalized harmonic number $H(n, p)$;

$$H(n, p) = \sum_{k=1}^{n} k^{-p}. \tag{69}$$

As the number of samples, increase, we expect M to be close to N. It is numerically more stable to consider the statistics of the gap G, defined as:

$$G = \frac{N - M}{N}. \tag{70}$$

Rewriting $E(M)$ for $E(G) = (N - E(M))/N$, we find

$$E(G) = \frac{1}{N^{S+1}}H(N-1, -S). \tag{71}$$

A similar calculation for the second moments results in

$$\text{Var}\, G = E((G - E(G))^2) \tag{72}$$

$$= E(G) \cdot (2 - E(G) - \frac{1}{N}) - \frac{2H(N-1, -(S+1))}{N^{S+2}}. \tag{73}$$

In the Max64 test, we are using large values of $N > 2^{40}$, so it is not feasible to use the explicit sum 69. Using the approximation,

$$\sum_{k=1}^{n} k^s \approx \frac{1}{s+1}(n+1/2)^{s+1} + \text{const} \tag{74}$$

we obtain

$$E(G) \approx g\,. \tag{75}$$

and

$$Var(G) \approx (2 - g - 1/N)g - 2h\,. \tag{76}$$

where

$$g = \frac{e^{-(S+1)*z}}{S+1}\,, \tag{77}$$

$$h = \frac{e^{-(S+2)*z}}{S+2}\,, \tag{78}$$

$$z = \frac{1}{2N}\left(1 + \frac{1}{4N}\right)\,. \tag{79}$$

Numerical simulations suggest this is a relative $O(1/N^2)$ approximation. With accurate moment results to compute the mean and standard deviation, we performed a set of tests using $S = 19$ samples and $N = 2^{63}$.

Table 3 Constitutes a numerical statistical proof of failure for some reference PRNGS in the Dieharder suite (except AES_OFB which is provided as a counter example). The impossible accumulated luck is proof either the model or the source of randomness is wrong. The last row contrasts this with the AES_OFB reference generator. Here, even after 1,000,000 trials, the accumulated luck is unsurprising, which is strong evidence the model is correct.

The testrng tool proves the first 22 generators are not random in about 20 seconds, while getting indeterminate results using the Dieharder suite took about 1 week of CPU time on a faster server.

Luck interpretation of dieharder results

The second approach is a more qualitative result. The typical output of a dieharder test is a p-value, which should be uniformly distributed if the test is correct and the generator is indistinguishible from random by the given test. Because there are many tests and many generators, weak/fail results are expected to happen. This is frustrating because it makes the results more difficult to interpret. By repeating a test, a consistently weak or failed test is a more reliable outcome. But that gives little concrete advice to determine how many failures constitute a real failure, nor does it allow for many-almost failures to have any meaning.

Generator	Trials	Outcome	Normal luck
borosh13	3,005	lucky	$\approx 1 - 10^{-45}$
rand	11,292	unlucky	$\approx 10^{-45}$
coveyou	898	unlucky	$\approx 10^{-45}$
knuthran	12,002	lucky	$\approx 1 - 10^{-45}$
ran3	3,552	lucky	$\approx 1 - 10^{-45}$
r250	95,415	lucky	$\approx 1 - 10^{-45}$
ranlux	503	lucky	$\approx 1 - 10^{-45}$
ranlux389	911	lucky	$\approx 1 - 10^{-45}$
ranlxs0	613	lucky	$\approx 1 - 10^{-45}$
ranlxs1	387	lucky	$\approx 1 - 10^{-45}$
ranlxs2	790	lucky	$\approx 1 - 10^{-45}$
random8-bsd	19,311	lucky	$\approx 1 - 10^{-45}$
random8-glibc2	1,914	unlucky	$\approx 10^{-45}$
ranmar	551	lucky	$\approx 1 - 10^{-45}$
slatec	487	lucky	$\approx 1 - 10^{-45}$
transputer	9,450	unlucky	$\approx 10^{-45}$
uni	100	lucky	$\approx 1 - 10^{-45}$
vax	20,523	lucky	$\approx 1 - 10^{-45}$
waterman14	2,512	unlucky	$\approx 10^{-45}$
zuf	271	lucky	$\approx 1 - 10^{-45}$
R_knuth_taocp	16,010	lucky	$\approx 1 - 10^{-45}$
R_knuth_taocp2	8,254	lucky	$\approx 1 - 10^{-45}$
AES_OFB	1,000,000	normal	0.2854

Table 3: Accumulated normal luck using the Max64 tests for failed reference Dieharder PRNGs and AES_OFB as strong counterexample.

To create a more summary proof-of-failure statistic, we ran 10 instances of each known-good test against each algorithmic PRNG, and assumed the p-value of each outcome was the outcome of a 1-dimensional normal distribution. This allowed for the computation of a z-score. In this way, for each random number generator, each p-value for all tests provided a normal luck $z_L = z - \sqrt{1/2}$ and $d_f = 1$. These were accumulated over all 590 tests for the following summary statistics.

Extreme $p = 1$ and $p = 0$ outcomes were assigned a z-score of ± 4, since and error of 10^{-6} is possible from the formatting of the results. Also, the rgb_minimum tests were ignored since they failed with $p = 0$ on every generator

Some notable entries are gfsr4, which failed 1 test but did not achieve a provably surprizing normal luck result. By contrast R_knuth_taocp2 failed no tests and had relatively few weak results, but had a much more surprizing z_L. Max64 proves the latter is too lucky to be random. The ranlxs* vs the random32-* entries show neither Dieharder nor Max64 are uniformly better at disproving the randomness of a generator.

If one dedicated a 100 weeks of cpu time to this, it would be interesting to compare a larger data set. There are many highly unlikely results in the luck summary which are not statistical proofs, however because of the cavalier approach to the extraction of the initial normal luck, the results always have to be made in comparision to a strong generator. On the other hand, if the Dieharder suite provided z_L and d_f values for each test, it would be a trival matter to provide a summative provable result.

There are other ways these results could be significantly improved. First, the actual statistic could be used to obtain a correct z_L score to combine. Second, the z_L score could be reported instead of inferring it from the p-value so that extreme results could be incorporated into the luck estimates with greater accuracy. Finally, once a fixed number of tests are computed in the suite, the tests should be repeated so long as the contribution to the overall normal luck is monotonic. For correctly random luck, this introduces a small number of additional runs, but consistently lucky/unlucky trending tests will culminate into a proof of failure. These results would be much easier to interpret overall.

PRNG	z_L	#WEAK	#FAIL	$\lvert z_L \rvert > 10$	Max64
borosh13	61.7	17	435	lucky	lucky
cmrg	2.4	19	0		
coveyou	59.6	6	418	lucky	unlucky
fishman18	2.8	8	0		
fishman20	5.6	14	10		
fishman2x	2.9	16	0		
gfsr4	4.1	17	1		
knuthran	2.4	14	0		
knuthran2	3.0	13	0		
lecuyer21	4.8	10	10		
minstd	4.5	15	10		
mrg	1.0	12	0		
mt19937	2.5	13	0		
mt19937_1999	3.5	18	0		
mt19937_1998	1.7	12	0		
r250	20.3	39	50	lucky	lucky
ran0	5.4	12	10		
ran1	1.9	7	0		
ran2	1.8	8	0		
ran3	50.5	9	316	lucky	lucky
rand	58.3	18	388	lucky	unlucky
rand48	5.1	17	10		
random128-bsd	6.6	19	10		
random128-glibc2	7.8	22	10		
random128-libc5	6.8	25	10		
random256-bsd	4.0	14	0		
random256-glibc2	2.7	11	0		
random256-libc5	3.9	22	0		
random32-bsd	54.1	36	323	lucky	
random32-glibc2	53.0	40	312	lucky	
random32-libc5	53.4	35	321	lucky	
random64-bsd	21.8	46	59	lucky	
random64-glibc2	21.4	42	57	lucky	
random64-libc5	22.3	43	63	lucky	

Table 4: Normal luck estimates using internal dieharder tests summarizing 590 tests assuming the p-value of each test came from a 1-dimensional normal distribution and compared against the Max64 test with the same $\lvert z_L \rvert > 10$ cutoff criteria.

PRNG	z_L	#WEAK	#FAIL	$\lvert z_L \rvert > 10$	Max64
random8-bsd	57.8	14	390	lucky	lucky
random8-glibc2	58.4	14	391	lucky	unlucky
random8-libc5	58.2	13	390	lucky	lucky
random-bsd	5.1	10	11		
random-glibc2	7.1	18	10		
random-libc5	5.6	19	10		
randu	67.8	27	497	lucky	unlucky
ranf	8.0	19	10		
ranlux	3.5	16	0		lucky
ranlux389	2.7	12	0		lucky
ranlxd1	0.9	10	0		
ranlxd2	1.3	8	0		
ranlxs0	2.3	8	0		lucky
ranlxs1	2.0	14	0		lucky
ranlxs2	2.3	10	0		lucky
ranmar	2.4	12	0		lucky
slatec	67.4	25	485	lucky	lucky
taus	1.4	10	0		
taus2	1.4	10	0		
taus113	1.8	7	0		
transputer	62.3	6	448	lucky	unlucky
tt800	1.0	9	0		
uni	12.9	10	40	lucky	lucky
uni32	10.8	12	31	lucky	normal
vax	58.0	19	384	lucky	lucky
waterman14	61.4	12	437	lucky	unlucky
zuf	3.3	13	0		lucky
ca	6.6	16	10		
uvag	3.0	14	0		
AES_OFB	1.5	8	0		
Threefish_OFB	3.2	20	0		
kiss	2.6	12	0		
superkiss	7.7	12	20		
R_wichmann_hill	2.1	12	0		
R_marsaglia_multic.	5.7	16	10		
R_super_duper	10.5	12	25	lucky	
R_mersenne_twister	2.1	14	0		
R_knuth_taocp	2.4	12	0		lucky
R_knuth_taocp2	5.1	25	0		lucky

Table 5: Table 5 continued.

Proofs

Luck more generally.

Fundamentally, defining luck depends on a partial ordering of the elements of the probability space in question. In a finite or countable space, this can be done just from the probability. In larger spaces more topological structure is required:

$$y < x \text{ iff } \lim_{r \downarrow 0} \frac{p(B_r(x))}{p(B_r(y))} < 1\,. \tag{80}$$

Here, $p(S)$ denotes the probability of an outcome in $S \subseteq X$, and $B_\varepsilon(x) \subseteq X$ is the open ball of radius r in the topology of X.

It is straightforward to prove the transitivity of (80). Using this ordering, luck can be generalized as follows:

$$\Omega(x) = \{y \mid y < x\}\,, \tag{81}$$

$$|\Omega(x)| = p(\Omega(x))\,, \tag{82}$$

$$\omega(x) = \{y \mid y \not< x \text{ and } x \not< y\}\,, \tag{83}$$

$$|\omega(x)| = p(\omega(x))\,, \tag{84}$$

and

$$L(x) = |\Omega(x)| + \frac{1}{2}|\omega(x)|\,. \tag{85}$$

From the introduction

Theorem 1. *Range of Luck. For any probability space,*

$$0 \leq L(x) \leq 1\,. \tag{86}$$

Proof. From the definition, $L(x) = |\Omega(x)| + \frac{1}{2}|\omega(x)|$, which is clearly non-negative, and $\Omega(x)$ and $\omega(x)$ are disjoint subsets of the probability space. Here the $|\cdot|$ notation is the measure of these sets in the probability space, and their union is at most the whole space, so $L(x) \leq |\Omega(x) \cup \omega(x)| \leq 1$.

Theorem 2. *Lucky values. If $L(x)$ is close to 1, then $p(x)$ is relatively small and most outcomes have a higher probability (you are lucky).*

Proof. Since $L(x)$ is close to 1, write $L(x) = 1 - \varepsilon(x)$, where $\varepsilon(x)$ is a small non-negative number.

First, since $\Omega(x)$ and $\omega(x)$ are disjoint subsets of the space of outcomes X, $|\Omega(x)| + |\omega(x)| \leq 1$, or

$$|\Omega(x)| \leq 1 - |\omega(x)| \,.$$

Second, $L(x) = |\Omega(x)| + \frac{1}{2}|\omega(x)|$, which by the first inequality can be bounded as $1 - \varepsilon(x) \leq 1 - |\omega(x)| + \frac{1}{2}|\omega(x)|$, which can be rearranged as

$$|\omega(x)| \leq 2\varepsilon(x) \,.$$

Third, $|\Omega(x)| + \frac{1}{2}|\omega(x)| = 1 - \varepsilon(x)$, or $|\Omega(x)| = 1 - \varepsilon(x) - \frac{1}{2}|\omega(x)|$, which by the second inequality,

$$|\Omega(x)| \geq 1 - 2\varepsilon(x) \,. \tag{87}$$

Thus at least a $1 - 2\varepsilon(x)$ fraction of the probability space have a higher probability of occurring.

Theorem 3. *Unlucky values. If $L(x)$ is close to 0, then $p(x)$ is comparatively large, and most outcomes would have a lower probability (you are unlucky).*

Proof. Since $L(x)$ is close to 0, write $L(x) = \varepsilon(x)$, where $\varepsilon(x)$ is a small non-negative number.

First, $L(x) = |\Omega(x)| + \frac{1}{2}|\omega(x)|$, so

$$\omega(x) \leq 2\varepsilon(x) \,.$$

Second, $|\Omega(x)| + |\omega(x)| = |\Omega(x)| + \frac{1}{2}|\omega(x)| + \frac{1}{2}|\omega(x)| \leq 2\varepsilon(x)$, so

$$|\Omega(x)| + |\omega(x)| \leq 2\varepsilon(x) \,.$$

Third, $|X - [\Omega(x) \cup \omega(x)]| \geq 1 - |\Omega(x) \cup \omega(x)| \geq 1 - 2\varepsilon(x)$, or

$$|X - [\Omega(x) \cup \omega(x)]| \geq 1 - 2\varepsilon(x) \,.$$

Thus at least a $1 - 2\varepsilon(x)$ fraction of the probability space have a lower probability of occurring.

Theorem 4. *On average, luck is always 50:50.*

$$E(L) = \frac{1}{2} \,.$$

Proof. This is an application of the next theorem where $f(L) = L$.

Theorem 5. *Smooth uniformity - finite space X.*

$E(f(L)) = \int_0^1 f(L)dL - \varepsilon$, *where* $|\varepsilon| \leq \max|f''| \cdot \max|\omega|^2/24$.

Since the definition of luck only depends on the probabilities of outcomes, it is natural to consider the set of equivalence classes $[x]$ *from* $[X]$ *defined by equal probabilities:* $[x] = \{y \mid p(y) = p(x)\}$.

We also use the midpoint integration estimate:

$$\int_a^b f(L)\,dL = (b-a)f(\frac{a+b}{2}) + \frac{(b-a)^3}{24}f''(\xi), \text{where } a < \xi < b.$$

$$\int_0^1 f(L)\,dL = \sum_{[x]} \int_{L([x])-\frac{1}{2}|\omega([x])|}^{L([x])+\frac{1}{2}|\omega([x])|} f(L)\,dL\,, \tag{88}$$

$$= \sum_{[x]} \left\{ |\omega([x])| f(L([x])) + \frac{|\omega([x])|^3}{24} f''(\xi_{[x]}) \right\}, \tag{89}$$

$$= \sum_x p(x) f(L(x)) - \varepsilon\,, \tag{90}$$

$$= E(f(L)) - \varepsilon\,, \tag{91}$$

$$\varepsilon = \sum_{[x]} \frac{|\omega([x])|^3}{24} f''(\xi_{[x]})\,, \tag{92}$$

$$|\varepsilon| \leq \frac{1}{24} \max_x |\omega(x)|^2 \cdot \max_{0 \leq L \leq 1} |f''(L)|\,. \tag{93}$$

Theorem 6. *Best of luck - finite space X. The fact that the definition of luck samples the intervals in theorem 5 at the centers. This is by design and is the only choice that leads to a second-order error error (*f''*) with the following constraints:*

- *Luck is constant for constant probabilities: if* $p(x) = p(y)$ *then* $L(x) = L(y)$.
- *Luck is increasing for decreasing probabilities: if* $p(x) < p(y)$ *then* $L(x) > L(y)$.

$$E(f(L)) = \sum_x p(x) f(L(x)) \tag{94}$$

using equality of luck for equal probabilities

$$= \sum_{[x]} |\omega([x])| f(L([x])) \tag{95}$$

Optimizing this sum as a quadrature of $\int_0^1 f(L)\,dL$ *is a classic question answered by the gauss points (centers of the intervals). The fact that* $L(x)$ *is increasing orders the gauss points to our definition of luck.*

Theorem 7. *Moments - finite space X. For $p \geq 2$, $E(L^p) = 1/(p+1) - \varepsilon$, $0 \leq \varepsilon \leq p \cdot (p-1) \max |\omega|^2/24$.*

Proof. This is an example of $f(L) = L^p$ with $p \geq 1$ in the theorem above, and noting that $f''(L)$ is non-negative, so ε in (92) must be non-negative.

Theorem 8. *Interval uniformity - finite space X. For $0 \leq a \leq b \leq 1$, $E(L \in [a,b]) = b - a - \varepsilon$, $|\varepsilon| \leq \max |\omega|$.*

Proof. Let $f(L)$ be the characteristic function for the closed interval $[a,b]$.

$$b - a = \int_0^1 f(L)\, dL \tag{96}$$

$$= \sum_{[x]} \int_{L([x]) - \frac{1}{2}|\omega([x])|}^{L([x]) + \frac{1}{2}|\omega([x])|} f(L)\, dL\,, \tag{97}$$

$$= E(f(L)) + \varepsilon\,, \tag{98}$$

$$\varepsilon = \sum_{[x]} \int_{L([x]) - \frac{1}{2}|\omega([x])|}^{L([x]) + \frac{1}{2}|\omega([x])|} (f(L) - f(L([x])))\, dL\,. \tag{99}$$

Now in the sum that defines ε, there are at most 2 discontinuities in an otherwise constant 0, 1 or -1 integrand. If none occur in an interval, that term is exactly zero. If one occurs, the integrand is zero for at least $\frac{1}{2}$ of the interval, so that error is at most $\frac{1}{2}|\omega([x])|$, and can be only once more in one other interval (with the same kind of error). And if 2 occur, they occur nowhere else and the error is bounded by $|\omega([x])|$. In each case we have our theorem.

From Normal

The approximation (39), i.e.,

$$L \approx L_1 = \frac{1}{2}\left[1 + \operatorname{erf}(|\sqrt{\Sigma^{-1}}(x - \mu)| - \sqrt{n - 1/2})\right]\,,$$

mostly comes from a Taylor expansion of the log of the integrand in (35). This in fact results in the approximation,

$$L \approx L_2 = \frac{1}{2}\left[1 + \operatorname{erf}(|\sqrt{\Sigma^{-1}}(x - \mu)| - \sqrt{n})\right]\,.$$

The shift from $\sqrt{n}$ to $\sqrt{n - 1/2}$ was a result of numerical experimentation, summarized in figure 15. It shows $\max |L - L_1|$ is has a $4\times$ smaller than $\max |L - L_2|$ for $4 \leq n \leq 100,000$.

From Computation

In this section, we again consider only a finite probability space X with $|X|$ elements x_k, $k = 1, \ldots, |X|$ and probabilities $p_k = p(x_k)$.

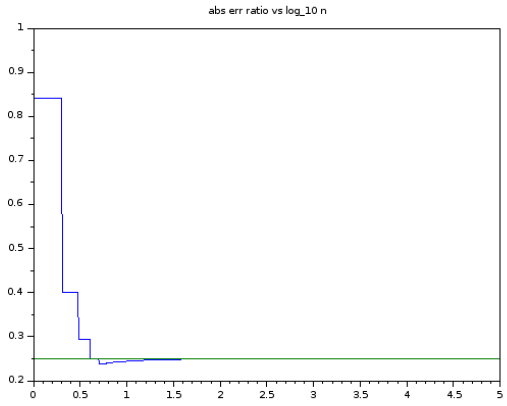

Figure 15: This is a plot of $E(n) = \max_{0\le R\le\infty} |L - L_1| / \max_{0\le R\le\infty} |L - L_2|$ vs $\log_{10} n$. The $\sqrt{n - 1/2}$-shift is always better than $\sqrt{n}$, and $4\times$ better for $n \ge 4$ ($\log_{10} 4 \approx 0.6$). The constant line is $1/4$ for reference.

Let $S = (s_1, \ldots, s_n)$ be $|S|$ independent samples taken from X. We define an estimator for $L(x)$ as:

$$\begin{aligned}\ell(x) =& \frac{1}{|S|} \{\#\text{ of outcomes in S more probable than } x\} \\ &+ \frac{1}{2|S|} \{\#\text{ of outcomes in S equally probable to } x\}\end{aligned}$$

We asserted that

$$E(\ell(x)) = L(x)\,. \tag{100}$$

$$E([\ell(x) - L(x)]^2) = \frac{1}{|S|}\left[L(x)\cdot(1 - L(x)) - \frac{1}{4}|\omega(x)|\right]\,. \tag{101}$$

For the remainder of this section, we consider $x \in X$ to be a fixed choice, and so drop the $\cdot(x)$ functional notation, which would otherwise pepper every equation. This way (100-101) becomes

$$E(\ell) = L\,. \tag{102}$$

$$E([\ell - L]^2) = \frac{1}{|S|}\left[L\cdot(1 - L) - \frac{1}{4}|\omega|\right] \tag{103}$$

The point of this section is to work out the details of these facts.

The definition of ℓ does not depend on the ordering of the sample S, so the vector of counts $(c_1, \ldots, c_{|X|})$ is equivalent information about the sample for our purposes. We define c as the vector of counts of elements in S, so that x_k occurs c_k times in the sample S.

The probability of obtaining a sample counts c is given by the multinomial distribution

$$P(c) = \binom{n}{c} p^c = \frac{n!}{c_1!\cdot\cdots\cdot c_{|X|}!} p_1^{c_1}\cdots p_{|X|}^{c_{|X|}}\,, \text{where } n = \sum c = |S|. \tag{104}$$

Here are some well-known moments of the multinomial distribution (the sum is over all samples with $\sum c = n$):

$$\sum_c P(c) = 1\,, \tag{105}$$

$$\sum_c P(c)c = np\,, \tag{106}$$

$$\sum_c P(c)cc^T = n\,\mathrm{diag}(p) + (n^2 - n)pp^T\,. \tag{107}$$

Here c and p are $|X| \times 1$ column vectors, and $\mathrm{diag}(p)$ is the diagonal $|X| \times |X|$ matrix with the probabilities p along the diagonal.

It is useful to introduce A and a, which are analogous to Ω and ω (recall that x is fixed in this discussion):

$$A = \{k \mid p(x_k) > p(x)\}\,, \tag{108}$$

$$= \{k \mid x_k \in \Omega\}\,, \tag{109}$$

$$|A| = \frac{1}{n}\sum_{k\in A} c_k\,. \tag{110}$$

Note that A is the index set of Ω, and so independent of c, but $|A|$ depends on the sample counts. By contrast, both Ω and $|\Omega|$ are constant (properties of the entire probability space) for a fixed choice of x.

Similarly,

$$a = \{k \mid p(x_k) = p(x)\}\,, \tag{111}$$

$$= \{k \mid x_k \in \omega\}\,, \tag{112}$$

$$|a| = \frac{1}{n}\sum_{k\in a} c_k\,. \tag{113}$$

Using the well-known multinomial moment results, we can compute some useful expected values of $|A|$ and $|a|$:

First $E(|A|) = |\Omega|$:

$$E(|A|) = \sum_c P(c)\frac{1}{n}\sum_{k\in A} c_k \tag{114}$$

$$= \frac{1}{n}\sum_{k\in A}\sum_c P(c)c_k \tag{115}$$

$$= \frac{1}{n}\sum_{k\in A} np_k \tag{116}$$

$$= |\Omega|\,. \tag{117}$$

Similarly $E(|a|) = |\omega|$.

Next $E(|A|^2) = \frac{1}{n}|\Omega|\cdot(1 - |\Omega|) + |\Omega|^2$:

Similarly, $E(|a|^2) = \frac{1}{n}|\omega|\cdot(1 - |\omega|) + |\omega|^2$.

$$E(|A|^2) = \frac{1}{n^2} \sum_{k,k' \in A} \sum_c P(c) c_k c_{k'} \,, \tag{118}$$

$$= \frac{1}{n^2} \sum_{k,k' \in A} \left[n \,\mathrm{diag}(p) + (n^2 + n) p p^T \right]_{k,k'} \,, \tag{119}$$

$$= \frac{1}{n} |\Omega| + \frac{n-1}{n} |\Omega|^2 \,, \tag{120}$$

$$= \frac{1}{n} |\Omega| \cdot (1 - |\Omega|) + |\Omega|^2 \,. \tag{121}$$

We also will need the cross term $E(|A||a|)$:

$$E(|A||a|) = \frac{1}{n^2} \sum_{k \in A, k' \in a} \sum_c P(c) c_k c_{k'} \,, \tag{122}$$

$$= \frac{1}{n^2} \sum_{k \in A, k' \in a} \left[n \,\mathrm{diag}(p) + (n^2 + n) p p^T \right]_{k,k'} \,, \tag{123}$$

$$= \frac{n-1}{n} |\Omega| |\omega| \,. \tag{124}$$

Note that the diagonal term $\mathrm{diag}(p)$ contributes nothing in this case because A and a are disjoint sets.

With these preliminaries,

$$\ell = |A| + \frac{1}{2} |a| \,, \tag{125}$$

And so

$$E(\ell) = |\Omega| + \frac{1}{2} |\omega| = L \,.$$

which is the first result.

For the variance,

$$E([\ell - L]^2) = E(\left[(|A| - |\Omega|) + \frac{1}{2} (|a| - |\omega|) \right]^2) \,, \tag{126}$$

$$= E([|A| - |\Omega|]^2) + E([|A| - |\Omega|][|a| - |\omega|]) + \frac{1}{4} E([|a| - |\omega|]^2) \,, \tag{127}$$

$$= E(|A|^2) - |\Omega|^2 + E(|A||a|) - |\Omega||\omega| + \frac{1}{4} \left\{ E(|a|^2) - |\omega|^2 \right\} \,, \tag{128}$$

$$= \frac{1}{n} \left\{ |\Omega|(1 - |\Omega|) - |\Omega||\omega| + \frac{1}{4} |\omega|(1 - |\omega|) \right\} \,. \tag{129}$$

A little algebra will show (129) is equal to

$$E([\ell - L]^2) = \frac{1}{n} \left\{ L \cdot (1 - L) - \frac{1}{4} |\omega| \right\} \leq \frac{L \cdot (1 - L)}{n} \,. \tag{130}$$

which is the other thing we wanted to prove.

Infinite and continuous spaces.

Theorems 4-6 have been proven for any finite probability space X.

$$T_4(X) : E(L(x)) = \frac{1}{2}\,, \tag{131}$$

$$T_5(X) : \left| E(f(L)) - \int_0^1 f(L)\, dL \right| \leq \frac{\max_L |f''(L)| \cdot \sup_x |\omega(x)|^2}{24}\,, \tag{132}$$

$$T_6(X) : 0 \leq \frac{1}{p+1} - E(L^p) \leq \frac{p \cdot (p-1) \cdot \sup_x |\omega(x)|^2}{24}\,. \tag{133}$$

But expectation $E(\cdot)$ and measure $|\cdot|$ are continuous in the space of probability measures, and the inequalities are closed, so by a cauchy sequence of measures, these statements are true in general.