WARREN D. MACEVOY

# LUCK

PERSONAL NOTES

PUBLISHED BY PERSONAL NOTES

WWW.COLORADOMESA.EDU



*First printing, October 2024*

# *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.

This may exist as another name, but I don't think so, and it really should be called luck. A Rose by the name of Rosaceae would be lost.

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 occasionally be surprised by the outcome.

For a continuous probability 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 probability $p(x) = \rho(x)dx$ 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 $x$:

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

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

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 $|\omega(x)| = 1$.

For the kind of distributions with small $\max|\omega|$, luck is a very uniformizing transformation (there can be no better, actually). If it were exactly uniform, the following would be correct with $\varepsilon = 0$. There is no other functional of the probability space alone with smaller error bounds:

- For any $f : [0,1] \to \mathbb{R}$ with bounded second derivative, $E(f(L)) = \int_0^1 f(L)dL - \varepsilon$, where $|\varepsilon| \leq \max|f''| \cdot \max|\omega|^2/24$.
- For $p \geq 2$, $E(L^p) = 1/(p+1) - \varepsilon$, $0 \leq \varepsilon \leq p \cdot (p-1)\max|\omega|^2/24$.
- For $0 \leq a \leq b \leq 1$, $E(L \in [a,b]) = b - a - \varepsilon$, $|\varepsilon| \leq \max|\omega|$.

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

**Example 1.** *Eight Fair 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.

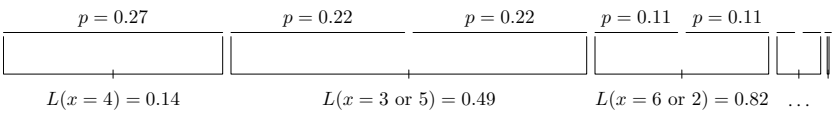



Figure 1: Luck to get $x$ heads on 8 fair coins. After the probabilities are arranged in decreasing order on the unit interval, $L(x)$ is the center point of equally probable outcomes.

*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 $0 \leq \varepsilon \leq \max|\omega|^2/12$.

For the coin distribution, the bound is $0 \leq \varepsilon \leq 0.016$, 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 $\frac{1}{2}$:*

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

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

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

**Example 2.** *Normal ($d_f = 1$). 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 $\mu$ and variance $\Sigma$ 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 $\mu$:*

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

*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*

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

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

*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.*

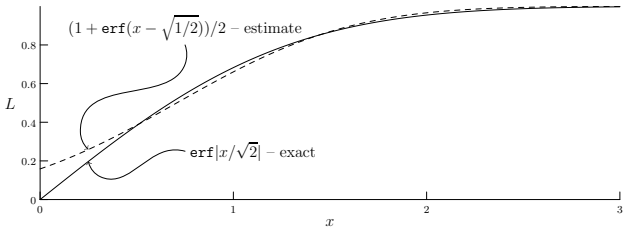



Figure 2: Exact (solid, equation 22) vs approximate (dashed, equation 23) luck for 1-d normal distribution.

**Example 3.** $\chi^2$ *p-value vs. L*

*For symmetric or monotonic one dimensional continuous distributions, the connection between luck and p-values can be pretty simple. However, the relation between p and L can be complex. Figure 11 shows the p-values vs luck for* $\chi^2_k(x)$ *for* $k = 4$ *distribution.*

*In the chapter on* $\chi^2$*, we show a good approximation for large k is*

$$L[\chi^2_k](x) \approx \mathrm{erf}\left|\sqrt{x} - \sqrt{k - 2}\right| \,. \tag{24}$$

## *Summary*

- Luck has a natural definition of

$$\begin{aligned}\mathrm{Luck}(x) = &\,\mathrm{Prob}(\text{anything more likely than } x) \\ &+ \frac{1}{2}\mathrm{Prob}(x \text{ or anything equally likely to } x) \,.\end{aligned}$$

- Luck ranges from 0 (no luck) to 1 (perfect luck).
- The expected value of luck is always exactly 1/2.

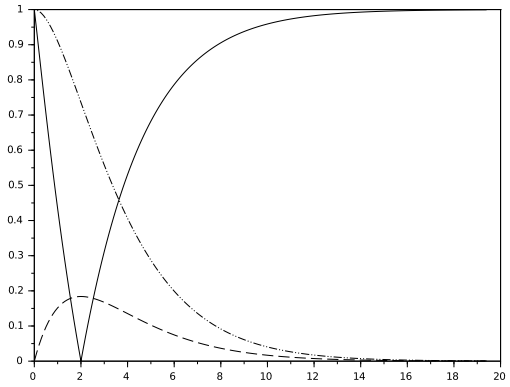



Figure 3: $p$-value (dotted-dashed) vs $L$ (solid) for the $\chi^2$ distribution for $k = 4$ (dashed). There is no simple relationship between $L$ and $p$.

- Luck is very nearly uniform, and there is no better map to a uniform distribution for a general probability distribution.
- For the $d_f = 1$ normal distribution with mean $\mu$ and standard deviation $\sigma$:
$$L[\mathcal{N}_{1d}(\cdot;\mu,\sigma^2)](x) = \text{erf}\left|\frac{x-\mu}{\sqrt{2}\sigma}\right| .$$
- For the $\chi^2_k$ distribution, there is no simple form, but for large $k$:
$$L[\chi^2_k](x) \approx \text{erf}\left|\sqrt{x} - \sqrt{k-2}\right| .$$

## *Exercises*

For the concept questions argue for discrete or finite probabilities only.

1. Can you show that no discrete probability has perfectly unlucky outcomes? That is, $L(x) = 0$ is impossible?
2. Can you show that it is impossible to observe perfect luck? That is $L(x) = 1$ implies $p(x) = 0$ for any discrete probabilty space. Hint: define $\Omega'(x) = \{y|p(y) < p(x)\}$ as the set of outcomes less likely than $x$, and note that $\{\Omega(x), \omega x, \Omega'(x)\}$ partition the probabilty space so $|\Omega(x)| + |\omega(x)| + |\Omega'(x)| = 1$.
3. Why is the expected value of luck always $\frac{1}{2}$?
4. What is the relationship between luck and probability?
5. Can you think of an example of an event that has a low probability but a high luck?

6. Can you think of an example of an event that has a high probability but a low luck?

7. What is the difference between luck and randomness?

8. How can understanding luck be useful in fields like statistics or decision making?

9. Can you think of a real-life scenario where considering luck could help you make a better decision?

10. Can you think of a situation where someone's perceived luck might be different from their actual luck?

11. Take the maximum of two fair coin flips (0 or 1), so $\text{Prob}(\max = 0) = 1/4$ and $\text{Prob}(\max = 1) = 3/4$. What is the luck of each outcome?

12. Take 3 independent trials like problem 11. Note there are $2^3$ outcomes but only 4 distinct probabilities. Show that

$$\begin{aligned} L((1,1,1)) &= \frac{27}{128} \approx 0.2109 \\ L((0,1,1)) = L((1,0,1)) = L((1,1,0)) &= \frac{81}{128} \approx 0.6328 \\ L((1,0,0)) = L((0,1,0)) = L((0,0,1)) &= \frac{117}{128} \approx 0.9141 \\ L((0,0,0)) &= \frac{127}{128} \approx 0.9922 \end{aligned}$$

13. Use a strong random number generator (like `random.org`) to make 3 independent trials for the maximum of 2 fair coin tosses. What is the luck of the outcome?

14. Now write down what feels like a random sequence of six zeros and ones — don't think too hard about it! Use these in pairs like problem 13. What was the luck of this outcome?

15. Calculate the luck $L(x)$ for the exponential distribution $p(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$ and $\lambda > 0$.

16. In *Rosencrantz and Guildenstern Are Dead*, Guildenstern tosses 85 heads in a row at the beginning of the play. Taken as a sequence of fair coin tosses, why is Guildenstern luck 1/2?

17. Ignoring the last coin toss, group the 84 heads as 42 experiment pairs like problem 11. Show from this perspective Guildenstern is unlucky:

$$L = \frac{109418989131512359209}{38685626227668133590597632} \approx 2.828 \times 10^{-6}\,.$$

18. Now consider the minimum-of-pairs problem for the 42 head-pairs. Show that for this problem Guildenstern is insanely lucky:

$$L = 1 - 2^{-85} \approx 1 - 2.585 \times 10^{-26}$$

## *Solutions*

1. Setting $L = 0$ in $L = |\Omega(x)| + \frac{1}{2}|\omega(x)|$ means $|\Omega(x)\| = 0$ and $|\Omega(x)\| = 0$, for a discrete probability this is impossible, since the probability this *and all more likely* outcomes is zero, then all the probabilities are zero. That is a contradiction since the probabilities should sum to 1.

2. Define $\Omega'(x)$ as set of outcomes with probablity smaller than $p(x)$, since this a partition of the probability space $||\Omega(x)| + |\omega(x)| + |\Omega'(x)| = 1$. Subtractiing the requirement that $|\Omega(x)| + \frac{1}{2}|\omega(x)| = 1$, we get $\frac{1}{2}|\omega(x)| + |\Omega'(x)| = 0$. Since both of these terms are non-negative, we have $|\omega(x)| = 0$ and $|\Omega'(x)| = 0$. So, for a discrete probability, the probability of the outcome must be exactly zero (a miracle). For continuous probability it must be measure zero, which is a mathematical way of saying a miracle as well. You should never observe a perfectly lucky outcome.

3. TBD.

4. Probaility is assigning a non-negative likelyhood $p(x)$ to each posssible outcome $x$ in a probability space. For example $x$ could be a password and $p(x)$ could be the probability that a given user will use that as a password the next time they are asked to create a password. Luck is a way to assign a number $L(x)$ to the outcome of choosing a password. It gives a concrete measure to the idea of how lucky ($L(x) \approx 1$) or unlucky ($L(x) \approx 0$) observing that result is.

5. Use $p(x) = 2^{-x}$, $x = 1, \ldots, \infty$ as a discrete probability space. As $x$ grows, the probablity decreases exponentially, but the related luck gets closer to 1: $L(x) = 1 - \frac{3}{2} \cdot 2^{-x}$.

6. This isn't really possible, since $L(x) \geq \frac{1}{2}|\omega(x)| \geq \frac{1}{2}p(x)$, it can't be smaller than $1/2$ of the probability itself. So the best you can do is to make this equal. For the previous example that is $p(1) = 1/2$ for which the luck is $1/4$. For a discrete distribution, the worst luck will be for the most likely outcome.

7. If a system is correctly described by some probabilities $p(x)$, the observed luck will look as uniformly random as possible.

8. Much of statistics are tools to help decide if a given probablity distribution correctly describes the system. Luck is a way to mathematically assesss that which fits reasonably well with the common notions of luck.

9. Luck is more intuitive than $p$-value, which is meaningless to anyone unfamiliar with statistics. This allows a statistician to convey a result more easily. It also identifies both outlying (lucky) or surprizingly conformant (unlucky) outcomes.

10. Uniform probability spaces are unintitive. For example, gettiing a 100 heads in a row is as equally likely as any other sequence. The mathematical luck of this case is $1/2$, while a human would view it as impossibly lucky. Looking at that sequence from a non-uniform space (maximums over pairs, for example) allows us to identify the degree of human surprize in this result.

11. The domain is $\{0,1\}$, the two possible values of the maximum, with $p(1) = \frac{3}{4}$ and $p(0) = \frac{1}{4}$, in decreasing order of probability. So $L(1) = /\frac{1}{2}p(1) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$, and $L(0) = p(1) + \frac{1}{2}p(0) = \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{4} = \frac{7}{8}$.

12. For a single trial, the probability for a maximum of $m$ is

$$p_1(m) = \begin{cases} 3/4 & \text{if } m = 1, \\ 1/4 & \text{if } m = 0 \end{cases} \tag{25}$$

The probabilities for three independent trials are given by product of likelyhoods of the outcome of each trial: $p_3(m_1, m_2, m_3) = p_1(m_1) \cdot p_1(m_2) \cdot p_1(m_3)$, the rest is applying definitions:

| $\lvert x\rvert$ | $p(x)$ | $\Omega(x)$ | $\lvert\Omega(x)\rvert$ | $\omega(x)$ | $\lvert\omega(x)\rvert$ | $L(x)$ |
|---|---|---|---|---|---|---|
| 0 | 0.4219 | {} | 0.0000 | {(1,1,1)} | 0.4219 | 0.2109 |
| 1 | 0.1406 | {(1,1,1)} | 0.4219 | {(0,1,1),(1,0,1),(0,0,1)} | 0.4219 | 0.6328 |
| 2 | 0.0469 | {(1,1,0),(1,0,1),(0,1,1)} | 0.8438 | {(1,0,0),(0,1,0),(0,0,1)} | 0.2188 | 0.8203 |
| 3 | 0.0156 | {(1,1,0),(1,0,1),(0,1,1),(1,0,0),(0,1,0),(0,0,1)} | 0.9844 | {(1,1,1)} | 0.0625 | 0.9609 |

Table 2: This is arranged in increasing luck (which is decreasing probability). Getting exactly $x = 4$ heads is unlucky, requiring only 1 = 14% luck, while getting $x = 0$ or $x = 8$ heads is almost 100% luck.

$$L((1,1,1)) = \frac{27}{128} \approx 0.2109,, \quad L((0,1,1)) = L((1,0,1)) = L((1,1,0)) = \frac{81}{128} \approx 0.6328,, L((1,0,0)) = L((0,1,0)) = L((0$$

13. The luck of the outcome depends on the specific outcome generated. However, on average, the luck of the outcome is $\frac{1}{2}$.

14. The luck of the outcome depends on the specific sequence generated. However, on average, the luck of the outcome is $\frac{1}{2}$.

15. The luck of the exponential distribution is given by:

$$L(x) = \int_x^\infty \lambda \exp(-\lambda y), dy \qquad = \exp(-\lambda x),,$$

so the luck of an outcome $x$ is $\exp(-\lambda x)$.

16. Taken as a sequence of fair coin tosses, the probability of getting 85 heads in a row is $(1/2)^{85}$, which is a very small probability. However, the probability of getting any specific sequence of 85 coin tosses is also $(1/2)^{85}$, so the luck of getting this specific sequence is neither lucky nor unlucky.

17. When the 84 heads are grouped into 42 pairs, each pair has a luck of $3/4$, since there are 3 out of 4 possible outcomes that are more likely than the outcome in each pair. Thus, the overall luck of this sequence is given by:

$$L = \left(\frac{3}{4}\right)^{42} \approx 2.828 \times 10^{-6},,$$

which is very unlucky.

18. The minimum-of-pairs problem for the 42 head-pairs is equivalent to the maximum-of-pairs problem for the 42 tail-pairs. Since the two problems are equivalent, their lucks must be the same. The luck of the maximum-of-pairs problem is given by:

$$L = 1 - \left(\frac{1}{2}\right)^{42} \approx 1 - 2.585 \times 10^{-26},,$$

which is very lucky.

# *Normal Distribution*

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

The $d_f = 1$ example in the introduction is just where where $\mu$ and $\Sigma$ are scalars.

$$P_{\text{normal}}(x; \mu, \Sigma) = \frac{e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}}{\sqrt{(2\pi)^{d_f} \det \Sigma}}, \tag{26}$$

where

$$\mu_i = E(x_i), \tag{27}$$

and

$$\Sigma_{ij} = E((x_i - \mu_i)(x_j - \mu_j)). \tag{28}$$

How lucky is some outcome $x$? From the definition:

$$L(x) = |\Omega(x)| + \frac{1}{2}|\omega(x)|, \tag{29}$$

where

$$\Omega(x) = \left\{ y \in \mathbb{R}^{d_f} \middle| P_{\text{normal}}(y) > P_{\text{normal}}(x) \right\} \tag{30}$$

$$= \left\{ y \in \mathbb{R}^{d_f} \middle| |\sqrt{\Sigma^{-1}}(y - \mu)| < |\sqrt{\Sigma^{-1}}(x - \mu)| \right\} \tag{31}$$

and

$$\omega(x) = \left\{ y \in \mathbb{R}^{d_f} \middle| |\sqrt{\Sigma^{-1}}(y - \mu)| = |\sqrt{\Sigma^{-1}}(x - \mu)| \right\}. \tag{32}$$

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

$$|\omega(x)| = \int_{\omega(x)} P_{\text{normal}}(y; \mu, \Sigma)\, dy = 0. \tag{33}$$

So

$$L(x) = |\Omega(x)| \tag{34}$$

$$= \int_{\Omega(x)} P_{\text{normal}}(y; \mu, \Sigma)\, dy. \tag{35}$$

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

$$L(x) = \int_{|z|<R} P_{\text{normal}}(z; 0, I)\, dz\,, \tag{36}$$

where $R = |\sqrt{\Sigma^{-1}}(x-\mu)|$. This can be evaluated in spherical coordinates:

$$L(x) = \frac{1}{\sqrt{(2\pi)^{d_f}}} \int_0^R \frac{d_f \pi^{d_f/2}}{\Gamma(\frac{d_f}{2}+1)} r^{d_f-1} e^{-\frac{1}{2}r^2}\, dr\,, \tag{37}$$

$$= \frac{\gamma(d_f/2, R^2/2)}{\Gamma(d_f/2)}\,. \tag{38}$$

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

$$\gamma(s,x) = \int_0^x t^{s-1} e^{-t}\, dt \tag{39}$$

$$\Gamma(s) = \gamma(s,\infty)\,. \tag{40}$$

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

$$L(x) \approx \frac{1}{2}\left[1 + \text{erf}\left(\left|\sqrt{\Sigma^{-1}}(x-\mu)\right| - \sqrt{d_f - 1/2}\right)\right]\,. \tag{41}$$

This result has a number of important consequences. The first is insight into the nature of statistics with many degrees of freedom. Our expectation before this calculation was that luck would start with $L = 0$ at $R = 0$ and quickly increase with $R$ before becoming exponentially close to 1 as $R$ became large. This is not at all the case:

> In large dimensions, normal observations are *not* crowded near $x \approx \mu$, but almost certainly (99.99%) within $\pm 3$ of the elliptical shell
>
> $$|\Sigma^{-1/2}(x-\mu)| = \sqrt{d_f - 1/2}\,.$$

Philosophically, this means nobody should be surprised about not precisely reaching their goals. If you have enough dimensions to your life to be interesting, you would be very unlucky to be close to the mark on each of them.

**Example 4.** *Large $d_f$ normal luck. We can also use (41) to disprove an observation came from a distribution. Suppose we have distribution parameters $\mu$ and $\Sigma$, and would like to know if they fit actual observations. A traditional approach requires a large sample to estimate $\mu$ and $\Sigma$, but we really 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 $10^{-15}$).*

$\Sigma$ 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)

R=**norm**(z)

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

This comes from a Taylor expansion of the log of the integrand in (37), and numerical experimentation on the 1/2 factor. The expansion is specifically invalid for $d_f = 1$, hence the difference between the general case and the $d_f = 1$ case (22).

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

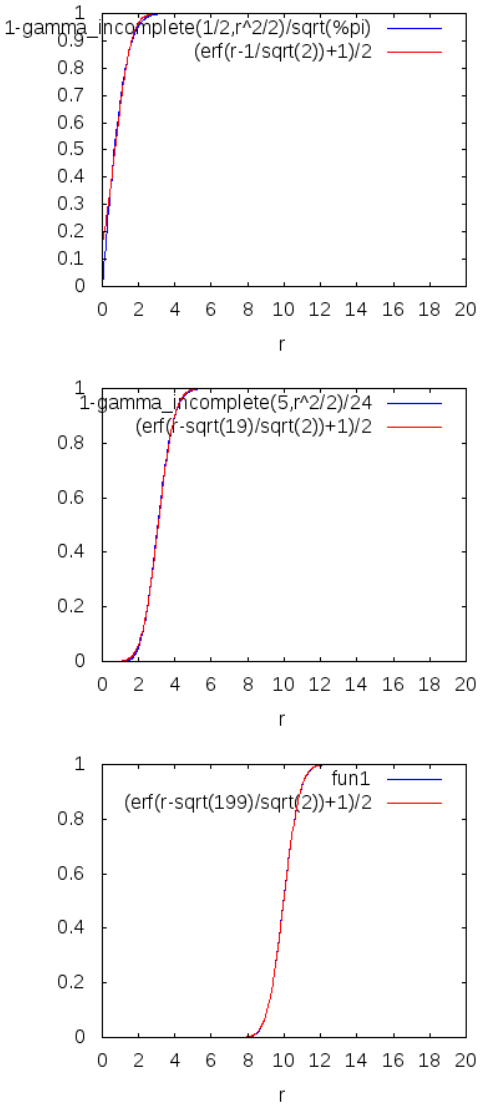



Figure 4: Exact (blue) vs approximate (red) luck for normal distribution for $d_f = 1$, 10, and 100.

| $L^{(x)}(x)$ | $L^{(y)}(x)$ | $L^{(x)}(y)$ | $L^{(y)}(y)$ |
|---|---|---|---|
| 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 3: 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 $100 \times 100$ 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.

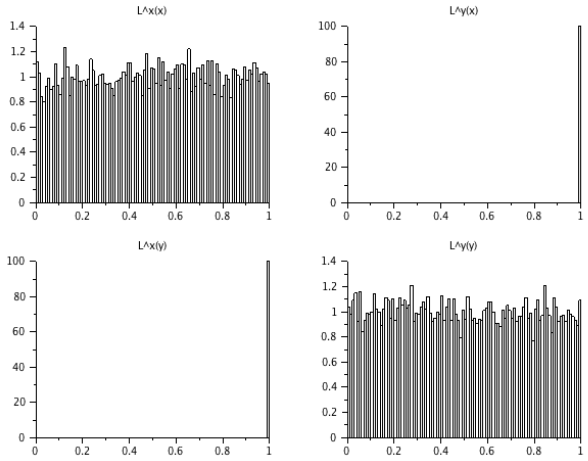



Figure 5: Histograms of 10,000 luck values in the same distributions as table 3. Upshot: $y$ values in the $x$ distribution are viewed as extremely lucky and vice-versa, while they are have uniform luck in their own respective distributions.

*Combining Normal Luck*

The approximate result (41) leads to a rule for combining (normal) luck: Suppose there are two independent normal distributions parametrized 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 $(x, y)$?

$$L(x,y) \approx \frac{1}{2}\left[1 + \operatorname{erf}(\sqrt{R_x(x)^2 + R_y(y)^2} - \sqrt{d_{fx} + d_{fy} - \frac{1}{2}})\right] , \quad (42)$$

where

$$R_x(x) = \operatorname{erf}^{-1}(2L_x - 1) - \sqrt{n_x - \frac{1}{2}} = \sqrt{\left(\Sigma^{(x)}\right)^{-1}}\left(x - \mu^{(x)}\right) , \quad (43)$$

$$R_y(y) = \operatorname{erf}^{-1}(2L_y - 1) - \sqrt{n_y - \frac{1}{2}} = \sqrt{\left(\Sigma^{(y)}\right)^{-1}}\left(y - \mu^{(y)}\right) . \quad (44)$$

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

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

$$z_L = \left|\sqrt{\Sigma^{-1}}(x-\mu)\right| - \sqrt{d_f - \frac{1}{2}} . \quad (45)$$

```
df=length(mu);
z=chol(Sigma)'\(x-mu);
zl=norm(z)-sqrt(df-1/2);
```

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

$$z_L^{A\times B} = \sqrt{\left(z_L^A + \sqrt{d_f^A - \frac{1}{2}}\right)^2 + \left(z_L^B + \sqrt{d_f^B - \frac{1}{2}}\right)^2} - \sqrt{d_f^A + d_f^B - \frac{1}{2}} , \quad (46)$$

and

$$d_f^{A\times B} = d_f^A + d_f^B \,. \tag{47}$$

Similarly, $n$ repeated independent experiments $A$ can be combined with

$$\begin{aligned} z_L^{A^n} =& \sqrt{\sum_{k=1}^{n}\left(z_L^{A_k} + \sqrt{d_f^A - \frac{1}{2}}\right)^2} \\ & - \sqrt{n\cdot d_f^A - \frac{1}{2}}\,, \end{aligned} \tag{48}$$

and

$$d_f^{A^n} = n d_f^A \,. \tag{49}$$

**Definition 5.** Normal luck $L_N$. If the combined dimension $d_f$ is large enough that the overall 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*,

$$L_N = \frac{1}{2}\left[1 + \mathrm{erf}(z_L)\right] \,. \tag{50}$$

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

$$L = \frac{\gamma(\frac{d_f}{2}, \frac{1}{2}(z_L + \sqrt{d_f - \frac{1}{2}})^2)}{\Gamma(d_f/2)} \,. \tag{51}$$

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

Combining $z$-luck is very useful for understanding the implications of multiple experiments. If the model is good, the combined $z$-luck value will to stay small (within $\pm 6$) 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 $\mu$ and/or $\Sigma$ 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 6-9 give Scilab functions for basic luck calculations for multivariate normal distributions.

```
function lnp=mnprobln(x,mu,Sigma)
  [df,nsamps]=size(x);
  sigma=chol(Sigma)';
  one=ones(1,nsamps);
  z=sigma\(x-mu*one);
  r2=sum(z.^2,'r');
  lnp=-r2/2-(df/2*log(2*%pi)+...
        sum(log(diag(sigma))));
endfunction
```

Figure 6: Scilab listing to compute the natural log of the probability of multivariate normal outcomes. $x$ is a $d_f \times n_{\text{samps}}$ matrix of outcomes, $\mu$ is a $d_f \times 1$ column vector of the mean, and $\Sigma$ is a $d_f \times d_f$ covariance matrix. The result `lnp` is a $1 \times n_{\text{samps}}$ row vector.

```
function zl=zluck(x,mu,Sigma)
  [df,nsamps]=size(x);
  sigma=chol(Sigma)';
  one=ones(1,nsamps);
  z=sigma\(x-mu*one);
  r=sqrt(sum(z.^2,'r'));
  zl=r-sqrt(df-1/2);
endfunction
```

Figure 7: Scilab listing to compute the $z$-luck. $x$ is a $d_f \times n_{\text{samps}}$ matrix of outcomes, $\mu$ is a $d_f \times 1$ column vector of the mean, and $\Sigma$ is a $d_f \times d_f$ covariance matrix. The result `zl` is a $1 \times n_{\text{samps}}$ row vector.

```
function [L,U]=mnluck(x,mu,Sigma)
  [df,nsamps]=size(x);
  one=ones(1,nsamps);
  zl=zluck(x,mu,Sigma);
  r2=(zl+sqrt(df/2)).^2;
  [L,U]=cdfgam("PQ",r2/2,(df/2)*one,one);
endfunction
```

Figure 8: Scilab listing to efficiently compute the luck of multivariate normal outcomes. $x$ is a $d_f \times n_{\text{samps}}$ of outcomes, $\mu$ is a $d_f \times 1$ column vector of the means, and $\Sigma$ is a $d_f \times d_f$ covariance matrix. The result $L$ is a $1 \times n_{\text{samps}}$ row vector of the luck associated with each outcome ($U = 1 - L$).

```
function [L,U]=normalluck(x,mu,Sigma)
  zl=zluck(x,mu,Sigma);
  L=0.5*erfc(-zl);
  U=0.5*erfc(zl);
endfunction
```

Figure 9: Like `mnluck` in figure 8, but approximates luck via (41).

## *Summary*

- For any distribution with finite mean $\mu$ and co-variance $\Sigma$, a useful parameter is the luck-adjusted $z$-score (45):

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

- For normal distributions (38),

$$L[\mathcal{N}(\cdot;\mu,\Sigma)](x) = \frac{\gamma(\frac{d_f}{2}, \frac{\left(z_L+\sqrt{d_f-\frac{1}{2}}\right)^2}{2})}{\Gamma\left(\frac{d_f}{2}\right)}\,.$$

- For approximately normal distributions (41),

$$L(x) \approx \frac{1}{2}\left[1+\text{erf}(z_L)\right]\,.$$

- Generalizing (46) and (48), for $n$ independent experiments $\{A_k\}_{k=1}^n$:

$$z_L^{A^n} = \sqrt{\sum_{k=1}^n \left(z_L^{A_k} + \sqrt{d_f^{A_k} - \frac{1}{2}}\right)^2} - \sqrt{\left(\sum_{k=1}^n d_f^{A_k}\right) - \frac{1}{2}}\,,$$

  and

$$d_f^{A^n} = \sum_{k=1}^n d_f^{A_k}\,.$$

- Seeing $|z_L| > Z$ should happen in less than $\frac{2Z}{\sqrt{\pi}}e^{-Z^2}$ fraction of cases. This can be be a statistical proof (likelihood $O(10^{-45})$) at $Z = 10$ with as little as one observation.

## *Exercises*

1. Take the maximum of two fair coin flips (0 or 1), so Prob(max $=$ 0) $=$ 1/4 and Prob(max $=$ 1) $=$ 3/4. What is the mean $\mu$ and standard deviation $\sigma$ of this distribution?

2. Take 10 independent trials like problem 1. Compute the value of $z$-luck for each trial.

3. Use (48) to combine the $z_L$-scores of problem 2. What is the overall normalized luck of the trials?

4. Now write down what feels like a random sequence of twenty zeros and ones — don't think too hard about it! Use these in pairs like problems 1-2. What was the normal luck of these trials?

5. Use a stats package to generate 10,000 values $x$ from a 1d normal distribution $A$ with mean of 1 and standard deviation of 1, and 10,000 values $y$ from a distribution $B$ with mean 1.1 and standard deviation 1.1. Combine the luck to get the overall luck of the 10,000 observations and show $y$ is lucky in $A$ and $x$ is unlucky in $B$.

# $\chi^2$ Distribution

The $\chi^2_k$ distribution is a non-symmetric distribution with support in the interval $(0, \infty)$, defined for $k = 1, 2, \ldots$ degrees of freedom as

$$\chi^2_k(x) = \frac{x^{k/2-1} e^{-\frac{x}{2}}}{2^{k/2}\Gamma(k/2)} \,. \tag{52}$$

The mean $\mu$ is $k$ and the variance $\sigma^2$ is $2k$.

The cumulative distribution of $\chi^2$ is surprisingly similar to luck for the $k$ degree-of-freedom normal distribution (37),

$$\mathrm{CDF}[\chi^2_k](x) = \int_0^x \chi^2_k(y)\, dy = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)} \,. \tag{53}$$

For this distribution, the $p$-value is usually the tail integral. That is,

$$p\text{-value} = 1 - \mathrm{CDF}[\chi^2_k](x) \,. \tag{54}$$

For $k = 1$ and $k = 2$ this distribution is monotonically decreasing, and so the luck is just the cumulative distribution function:

$$L[\chi^2_k](x) = \mathrm{CDF}[\chi^2_k](x), \text{ when } k = 1 \text{ or } k = 2. \tag{55}$$

However, for $k > 2$, the distribution is continuous with a single maximum at $x = k - 2$ and is otherwise monotonic. So the luck of an outcome is the integral of the PDF (52) between the observation $x$ and what we call the conjugate point $x^*$ where the distribution values are equal. This is illustrated in Figure 10 for $k = 4$ and $x = 5$.

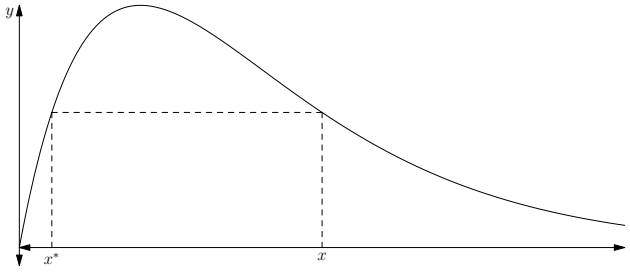



Figure 10: Illustration of the $x^*$ conjugate for the $\chi^2_4$ distribution at $x = 5$. The conjugate point has the same probability as the original: $\chi^2_4(x) = \chi^2_4(x^*)$. For $x = 5$, $x^* \approx 0.5368$.

**Definition 6.** $x^*$ for $\chi^2_k$. It is useful to generalize this a little to account for the $k = 1$ and $k = 2$ cases:

$$x^* = \begin{cases} \text{Point } y \neq x \text{ where } \chi^2_k(y) = \chi^2_k(x), & \text{if } k > 2 \text{ and } x \neq k-2, \\ k-2\,, & \text{if } k > 2 \text{ and } x = k-2, \\ 0\,, & \text{if } k \in \{1,2\}. \end{cases} \tag{56}$$

**Definition 7.** Luck $L$ for $\chi^2_k$ With the conjugate point $x^*$ defined this way, it is easy to express $\chi^2$-luck:

$$L(x) = \left|\text{CDF}[\chi^2_k](x) - \text{CDF}[\chi^2_k](x^*)\right|\,. \tag{57}$$

There is no algebraic solution for $x^*$, but $\chi^2_k$ is a smooth curve and so a Newton iteration works well on the logarithmic problem:

$$\log \chi^2_k(x^*) - \log \chi^2_k(x) = 0\,. \tag{58}$$

A suitable initial guess is to reflect the $x$ around the maximum point, which is asymptotically correct for points near the maximum, but then bound the root from below using the leading order approximation for small $x$,

$$x^*_{\text{reflect}} = \max(0, 2\sqrt{k-2} - \sqrt{x})^2\,. \tag{59}$$

$$x^*_{\text{lower\_bound}} = \left[\chi^2_k(x) 2^{k/2}\Gamma(k/2)\right]^{\frac{1}{k/2-1}} \tag{60}$$

and

$$x^*_0 = \max(x^*_{\text{reflect}}, x^*_{\text{lower\_bound}})\,. \tag{61}$$

For $x \neq k-2$, the root can quickly be found using the Newton iteration:

$$x^*_{n+1} = x^*_n + \frac{2x^*_n}{x^*_n - (k-2)}\left[\log \chi^2_k(x^*_n) - \log \chi^2(x)\right]\,. \tag{62}$$

Finally $x^* = \lim_{n\to\infty} x^*_n$ when $k > 2$ and $x \neq k-2$.

A plot of luck $L$ vs $p$-value for $\chi^2_4$ is given in Figure 11.

## *Approximating $L[\chi^2_k]$.*

A surprisingly good approximation for $L[\chi^2_k]$ comes from the following observations:

- Luck is zero at the maximum $x = k-2$.
- The tail of the distribution is $O(\text{erf}(\sqrt{x}))$.

This suggests

$$L[\chi^2_k](x) \approx \text{erf}\left|\sqrt{x} - \sqrt{k-2}\right|\,. \tag{63}$$

A plot of this estimate vs the numerical solution for large $k$ is given in Figure 12.

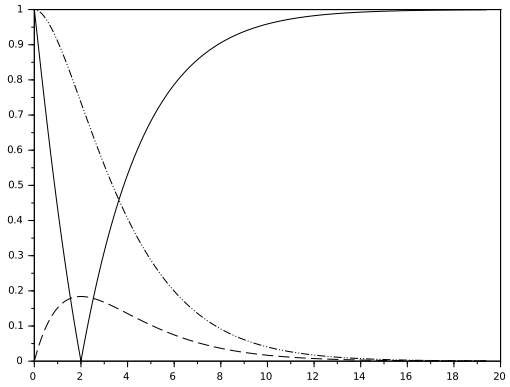



Figure 11: $p$-value (dotted) vs $L$ (solid) for the $\chi^2$ distribution for $k = 4$ (dashed). There is no simple relationship between $L$ and $p$.

## *Scilab reference code*

In order to avoid division by zero, we need a utility function which maps values away from zero defined by

$$\text{nonzero}_\varepsilon(x) = \begin{cases} x & \text{if } x \notin [-\varepsilon, \varepsilon]\,, \\ -\varepsilon & \text{if } x \in [-\varepsilon, 0)\,, \\ \varepsilon & \text{if } x \in [0, \varepsilon]\,. \end{cases} \tag{64}$$

Figure 13 gives a listing in scilab.

## *Summary*

- $\chi^2_k$ is a common non-symmetric distribution (52):

$$\chi^2_k(x) = \frac{x^{k/2-1}e^{-\frac{x}{2}}}{2^{k/2}\Gamma(k/2)}\,.$$

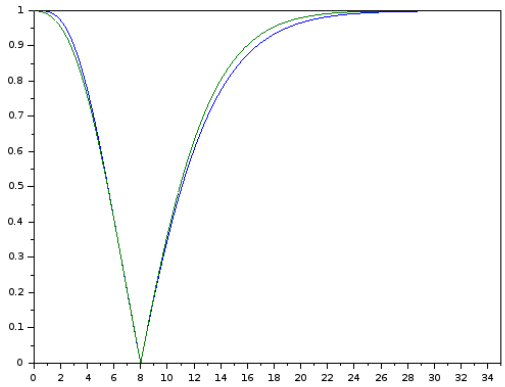
- For $k = 1$ or $k = 2$ degrees of freedom $\chi^2_k(x)$ is monotonic. But for $k > 2$ it has a single maximum at $x = k - 2$.
- We define $x^*$ as the conjugate point to $x$:

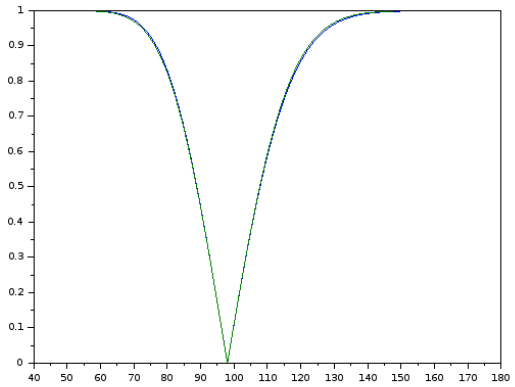
$$x^* = \begin{cases} \text{Solution } x^* \neq x \text{ to } \chi^2_k(x^*) = \chi^2_k(x) & \text{if } k > 2 \text{ and } x \neq k-2, \\ k-2 & \text{if } k > 2 \text{ and } x = k-2, \\ 0 & \text{if } k \in \{1, 2\}. \end{cases}$$

  This can be found using standard numerical techniques.

- With the conjugate point,

$$L[\chi^2_k](x) = \left| CDF[\chi^2_k](x^*) - CDF[\chi^2_k](x) \right|$$



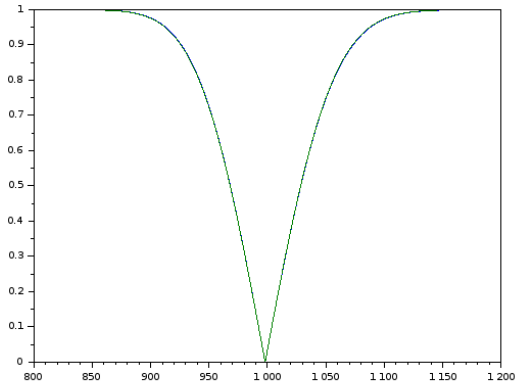




Figure 12: $L$ vs *Lestimate* for $k = 10$, 100, and $1,000$.

```
function y=nonzero(x,eps)
  y=x .* bool2s((x < -eps) | (eps < x)) + ...
    -eps*bool2s((-eps <= x) & (x < 0))+  ...
     eps*bool2s((0 <= x) & (x <= eps));
endfunction
```

Figure 13: Scilab utility function to avoid zero

```
function lnp=chi2probln(x,k)
  lnx=log(nonzero(x,number_properties('tiny')));
  lnp=lnx.*(k/2-1)-x./2-((k/2)*log(2)+gammaln(k/2));
endfunction
```

Figure 14: Scilab function to compute $\log \chi^2_k(x)$

```
function [P,Q]=chi2cdf(x,k)
  [dim,nsamps]=size(x);
  one=ones(1,nsamps);
  [P,Q]=cdfgam("PQ",x./2,k/2*one,one);
endfunction
```

Figure 15: Scilab function to compute $P = \mathrm{CDF}[\chi^2_k](x)$ and $Q = 1 - P$

```
function y=chi2conj(x,k)
  [df,nsamps]=size(x);
  lnp=chi2probln(x,k);
  yo=exp((1/(k/2-1))*(lnp+(k/2)*log(2)+gammaln(k/2)));
  yo=max(yo,max(0,2*sqrt(k-2)-sqrt(x)) .^ 2);
  y=yo;
  iterate=%T;
  i=0;
  cutoff=sqrt(%eps);
  while (iterate)
    z=y;
    df=chi2probln(z,k)-lnp;
    dln=2*z ./ nonzero(z-k+2,cutoff);
    y = z +  dln .* df;
    i=i+1;
    iterate = (i<1000) & or(abs(df)>cutoff);
  end
endfunction
```

Figure 16: Scilab function to compute $x^*$ for the $\chi^2_k$ distribution.

```
function [L,U]=chi2luck(x,k)
  y=chi2conj(x,k);
  [a,b]=chi2cdf(x,k);
  [A,B]=chi2cdf(y,k);
  L=abs(B-b);
  U=(A+b).*bool2s(x>=y)+(a+B).*bool2s(x<y);
endfunction
```

Figure 17: Scilab function to compute $L = L[\chi^2_k](x)$ and $U = 1 - L$.

- For large $k$ (63),

$$L[\chi^2_k](x) \approx \operatorname{erf}\left|\sqrt{x} - \sqrt{k-2}\right| .$$

## *Exercises*

1. Build an implementation of `chi2conj(x,k)` in your favorite stats toolbox.

2. An unfair die with un-normalized probabilities $(8,9,9,9,9,10)$ for face 1 through 6 is rolled 45 times giving the following counts: $(7,5,7,8,6,12)$. Show the $p$-value for this is 0.5591 and the luck for this is 0.7422 when comparing against expectations for a fair die.

3. A broken random generator produces $1,\ldots,6$ then repeats as a "random" die roll, so after 45 rolls, the counts are $(8,8,8,7,7,7)$. Show here the $p$-value is 0.9991 (meaning nothing interesting for a typical $\chi^2$ test) but the luck is 0.9721, which is surprisingly lucky.

# *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:

$$P_{\text{multinomial}}(x; p) = N! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!} \,. \tag{65}$$

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:

$$P_{\text{poisson}}(x; p) = \prod_{i=1}^{N_P} \frac{e^{-p_i} p_i^{x_i}}{x_i!} \,. \tag{66}$$

### *Approximating the Multinomial as a Normal Distribution*

The multinomial distribution can be approximated by a normal distribution,

$$\mu_i = N p_i \,, \tag{67}$$

$$\Sigma_{ij} = \begin{cases} N p_i (1 - p_i) & \text{if } i = j \,, \\ -N p_i p_j & \text{if } i \neq j \,. \end{cases} \tag{68}$$

But the covariance matrix $\Sigma$ 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

$$P(x) = \frac{\exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - N p_i)^2}{N p_i}\right)}{\sqrt{(2\pi)^{(n-1)} n \prod_{i=1}^{n} p_i}} \,. \tag{69}$$

subject to the constraint,

$$\sum_{i=1}^{n} x_i = N \,. \tag{70}$$

Changing coordinates to

$$z_i = \frac{x_i - Np_i}{\sqrt{Np_i}} \tag{71}$$

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

$$n_i = \sqrt{p_i} \tag{72}$$

This has the same form as the general Gaussian distribution, except it is one lower dimension, yielding,

$$L(x) = \frac{\gamma((n-1)/2, |z|^2/2)}{\Gamma((n-1)/2)}\,. \tag{73}$$

Comparing (73) with the normal distribution result (38), the $z$-luck value should be

$$z_L = \sqrt{\sum_{i=1}^{n} \frac{(x_i - Np_i)^2}{Np_i}} - \sqrt{n - \frac{3}{2}}\,, \tag{74}$$

$$d_f = n - 1\,. \tag{75}$$

## *Summary*

- The probability of obtaining $\{x_i\}_{i=1}^n$ counts after $N$ samples from $n$ bins with probabilities $\{p_i\}_{i=1}^n$ is given by the multinomial distribution,

$$P_{\text{multinomial}}(x; p) = N! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!}\,. \tag{76}$$

- There is no generally simple closed form to compute the luck associated with this distribution (see the next section on computation).

- Because of the constraint $\sum x_i = N$, one degree of freedom is lost, resulting in

$$z_L = \sqrt{\sum_{i=1}^{n} \frac{(x_i - Np_i)^2}{Np_i}} - \sqrt{n - \frac{3}{2}}\,, \tag{77}$$

$$d_f = n - 1\,. \tag{78}$$

# *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

$$E(\ell(x)) = L(x)\,, \tag{79}$$

and

$$E((\ell(x) - L(x))^2) \leq \frac{1}{n}L(x)\cdot(1 - L(x)) \leq \frac{1}{4n}\,. \tag{80}$$

**Example 5.** *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 18):*

$$y = \mathit{mulprobln(x,p)} = \log\left[\left(\sum_{i=1}^{n_p} x_i\right)!\prod_{i=1}^{n_p}\frac{p_i^{x_i}}{x_i!}\right]\,. \tag{81}$$

*Numerical estimates of luck requires samples from the given probability distribution. Figure 19 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 20 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 18: 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 19: Scilab listing to get a sample of outcomes from a multinomial distribution. $n_{\text{samps}}$ is the number of desired samples, $n_{\text{trials}}$ 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 20: 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 21: 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 22: 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 23: 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`.

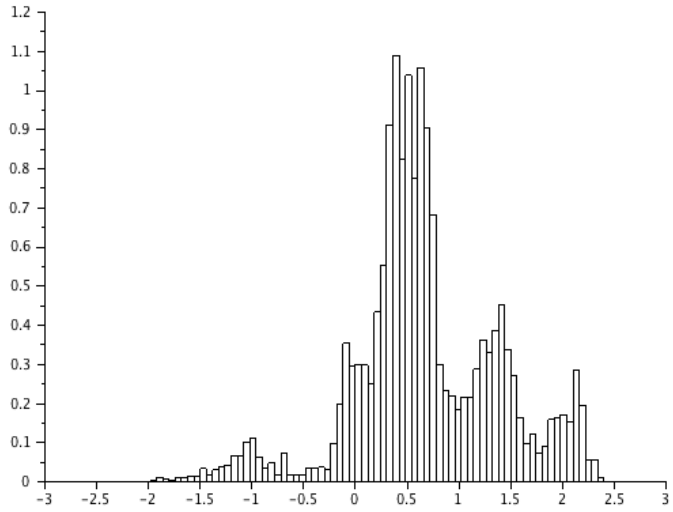



Figure 24: Histogram of z error values for 10,000 numerical approximations of luck. The maximum value of `sd` was `max(sd)=0.005`.

# *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 $\varepsilon$ 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 (45) of individual tests to obtain an overall normal luck score via the repeated use of (46). 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

$$P(M) = \text{Prob}(\max_{k=1}^{S} x_k = M) = \left(\frac{M}{N}\right)^S \cdot \left[1 - \left(\frac{M-1}{M}\right)^S\right] \tag{82}$$

$$= \frac{1}{N^S}\left(M^S - (M-1)^S\right). \tag{83}$$

The expected maximum value is given by

$$E(M) = \sum_{M=1}^{N} P(M)M, \tag{84}$$

$$= \frac{1}{N^S}\sum_{M=1}^{N}\left[M^{S+1} - (M-1)^S \cdot ((M-1)+1)\right], \tag{85}$$

$$= \frac{1}{N^S}\left[N^{S+1} - \sum_{M=1}^{N}(M-1)^S\right], \tag{86}$$

$$= N - \sum_{M=1}^{N-1}\left(\frac{M}{N}\right)^S, \tag{87}$$

$$= N - \frac{1}{N^S}H(N-1, -S). \tag{88}$$

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

$$H(n, p) = \sum_{k=1}^{n} k^{-p}. \tag{89}$$

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:

$$G = \frac{N - M}{N}. \tag{90}$$

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

$$E(G) = \frac{1}{N^{S+1}}H(N-1, -S). \tag{91}$$

A similar calculation for the second moments results in

$$\text{Var}\, G = E((G - E(G))^2) \tag{92}$$

$$= E(G) \cdot (2 - E(G) - \frac{1}{N}) - \frac{2H(N-1, -(S+1))}{N^{S+2}}. \tag{93}$$

In the Max64 test, we are using large values of $N > 2^{40}$, so it is not feasible to use the explicit sum 89. Using the approximation,

$$\sum_{k=1}^{n} k^s \approx \frac{1}{s+1}(n+1/2)^{s+1} + \text{const} \tag{94}$$

we obtain

$$E(G) \approx g\,. \tag{95}$$

and

$$Var(G) \approx (2 - g - 1/N)g - 2h\,. \tag{96}$$

where

$$g = \frac{e^{-(S+1)*z}}{S+1}\,, \tag{97}$$

$$h = \frac{e^{-(S+2)*z}}{S+2}\,, \tag{98}$$

$$z = \frac{1}{2N}\left(1 + \frac{1}{4N}\right)\,. \tag{99}$$

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 4 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 indistinguishable 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-fail results 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 4: 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 $\pm 4$, 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 surprising normal luck result. By contrast R_knuth_taocp2 failed no tests and had relatively few weak results, but had a much more surprising $z_L$. 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 comparison 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 trivial 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 | $\|z_L\| > 10$ | 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 5: 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 $|z_L| > 10$ 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 6: Table 6 continued.

# *Conclusions*

So what is luck? It is a mathematical definition with a useful analogy. That's important, because it makes it easier to understand "luck" over "$p$-value", or (worse) "generalized $p$-value".

Luck and generalized $p$-value are related by $L = 1 - p$ for probability spaces where $\max |\omega| = 0$.

The luck perspective on normal distributions provides surprising insights on the way observations accumulate and how to accommodate multiple observations into an overall statistic. There are also some handy algebraic tools to work out luck in the large.

We invoke an unproven generalization of the central limit theorem that evidently works.

Most generally, it is common to have only observations and perhaps a model for a statistical process. The numerical section describes how to estimate luck for such systems, and what kind of error tolerances to expect when using them. This important because so much of probability and statistics is non-parametric, including Bayesian statistics.

The chapter on randomness combines these ideas into a practical luck-based litmus test for randomness and demonstrates how it can be very efficient at identifying a misbehaved random process. It also shows how luck can be useful for summarizing a group of statistical tests. It also nicely illustrates a statistical computational proof-of-failure.

The particular tools really come from a perspective; luck as a useful point of view. It is more unifying and intuitive than a $p$-value. It has best-in-class properties compared to other $p$-value-like statistics.

Go out and try your luck today!

# *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:

$$y < x \text{ iff } \lim_{r \downarrow 0} \frac{p(B_r(x))}{p(B_r(y))} < 1\,. \tag{100}$$

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 (100). Using this ordering, luck can be generalized as follows:

$$\Omega(x) = \{y \mid y < x\}\,, \tag{101}$$
$$|\Omega(x)| = p(\Omega(x))\,, \tag{102}$$
$$\omega(x) = \{y \mid y \not< x \text{ and } x \not< y\}\,, \tag{103}$$
$$|\omega(x)| = p(\omega(x))\,, \tag{104}$$

and

$$L(x) = |\Omega(x)| + \frac{1}{2}|\omega(x)|\,. \tag{105}$$

## *From the introduction*

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

$$0 \leq L(x) \leq 1\,. \tag{106}$$

*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,*

$$|\Omega(x)| \geq 1 - 2\varepsilon(x) . \tag{107}$$

*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:*

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

$$\int_0^1 f(L)\,dL = \sum_{[x]} \int_{L([x])-\frac{1}{2}|\omega([x])|}^{L([x])+\frac{1}{2}|\omega([x])|} f(L)\,dL\,, \tag{108}$$

$$= \sum_{[x]} \left\{ |\omega([x])| f(L([x])) + \frac{|\omega([x])|^3}{24} f''(\xi_{[x]}) \right\}\,, \tag{109}$$

$$= \sum_x p(x) f(L(x)) - \varepsilon\,, \tag{110}$$

$$= E(f(L)) - \varepsilon\,, \tag{111}$$

$$\varepsilon = \sum_{[x]} \frac{|\omega([x])|^3}{24} f''(\xi_{[x]})\,, \tag{112}$$

$$|\varepsilon| \leq \frac{1}{24} \max_x |\omega(x)|^2 \cdot \max_{0 \leq L \leq 1} |f''(L)|\,. \tag{113}$$

**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 $L(x) > L(y)$.*

$$E(f(L)) = \sum_x p(x) f(L(x)) \tag{114}$$

*using equality of luck for equal probabilities*

$$= \sum_{[x]} |\omega([x])| f(L([x])) \tag{115}$$

*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 $\varepsilon$ in (112) 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* $[a,b]$.

$$b - a = \int_0^1 f(L)\, dL \tag{116}$$

$$= \sum_{[x]} \int_{L([x])-\frac{1}{2}|\omega([x])|}^{L([x])+\frac{1}{2}|\omega([x])|} f(L)\, dL\,, \tag{117}$$

$$= E(f(L)) + \varepsilon\,, \tag{118}$$

$$\varepsilon = \sum_{[x]} \int_{L([x])-\frac{1}{2}|\omega([x])|}^{L([x])+\frac{1}{2}|\omega([x])|} (f(L) - f(L([x])))\, dL\,. \tag{119}$$

*Now in the sum that defines* $\varepsilon$*, 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 (41), 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 (37). 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 25. 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 $p_k = p(x_k)$.

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

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

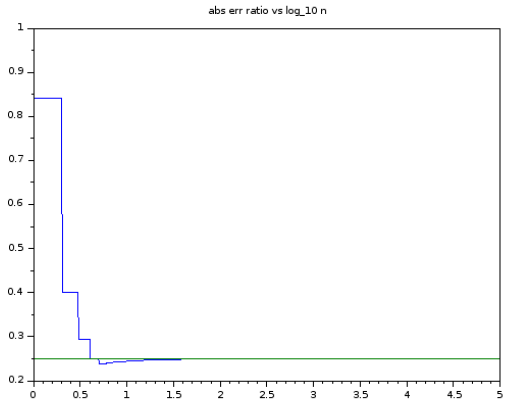



Figure 25: This is a plot of $E(n) = \max_{0\le R\le\infty} |L - L_1| / \max_{0\le R\le\infty} |L - L_2|$ vs $\log_{10} n$. The $\sqrt{n - 1/2}$-shift is always better than $\sqrt{n}$, and $4\times$ better for $n \ge 4$ ($\log_{10} 4 \approx 0.6$). The constant line is $1/4$ for reference.

We asserted that

$$E(\ell(x)) = L(x)\,. \tag{120}$$

$$E([\ell(x) - L(x)]^2) = \frac{1}{|S|}\left[L(x)\cdot(1 - L(x)) - \frac{1}{4}|\omega(x)|\right]\,. \tag{121}$$

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 (120-121) becomes

$$E(\ell) = L\,. \tag{122}$$

$$E([\ell - L]^2) = \frac{1}{|S|}\left[L\cdot(1 - L) - \frac{1}{4}|\omega|\right] \tag{123}$$

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

The definition of $\ell$ 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

$$P(c) = \binom{n}{c} p^c = \frac{n!}{c_1!\cdot\cdots\cdot c_{|X|}!} p_1^{c_1}\cdots p_{|X|}^{c_{|X|}}\,, \text{where } n = \sum c = |S|. \tag{124}$$

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

$$\sum_c P(c) = 1\,, \tag{125}$$

$$\sum_c P(c)c = np\,, \tag{126}$$

$$\sum_c P(c)cc^T = n\,\mathrm{diag}(p) + (n^2 - n)pp^T\,. \tag{127}$$

Here $c$ and $p$ are $|X| \times 1$ column vectors, and $\text{diag}(p)$ 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 $\Omega$ and $\omega$ (recall that $x$ is fixed in this discussion):

$$A = \{k \mid p(x_k) > p(x)\}\,, \tag{128}$$
$$= \{k \mid x_k \in \Omega\}\,, \tag{129}$$
$$|A| = \frac{1}{n}\sum_{k\in A} c_k\,. \tag{130}$$

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

Similarly,

$$a = \{k \mid p(x_k) = p(x)\}\,, \tag{131}$$
$$= \{k \mid x_k \in \omega\}\,, \tag{132}$$
$$|a| = \frac{1}{n}\sum_{k\in a} c_k\,. \tag{133}$$

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

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

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

$$E(|A|) = \sum_c P(c)\frac{1}{n}\sum_{k\in A} c_k \tag{134}$$
$$= \frac{1}{n}\sum_{k\in A}\sum_c P(c)c_k \tag{135}$$
$$= \frac{1}{n}\sum_{k\in A} np_k \tag{136}$$
$$= |\Omega|\,. \tag{137}$$

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

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

$$E(|A|^2) = \frac{1}{n^2}\sum_{k,k'\in A}\sum_c P(c)c_k c_{k'}\,, \tag{138}$$
$$= \frac{1}{n^2}\sum_{k,k'\in A}\left[n\,\text{diag}(p) + (n^2+n)pp^T\right]_{k,k'}\,, \tag{139}$$
$$= \frac{1}{n}|\Omega| + \frac{n-1}{n}|\Omega|^2\,, \tag{140}$$
$$= \frac{1}{n}|\Omega|\cdot(1-|\Omega|) + |\Omega|^2\,. \tag{141}$$

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

$$E(|A||a|) = \frac{1}{n^2} \sum_{k\in A,k'\in a} \sum_c P(c) c_k c_{k'} \,, \tag{142}$$

$$= \frac{1}{n^2} \sum_{k\in A,k'\in a} \left[ n \,\mathrm{diag}(p) + (n^2+n) pp^T \right]_{k,k'} \,, \tag{143}$$

$$= \frac{n-1}{n} |\Omega||\omega| \,. \tag{144}$$

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

With these preliminaries,

$$\ell = |A| + \frac{1}{2}|a| \,, \tag{145}$$

And so

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

which is the first result.

For the variance,

$$E([\ell - L]^2) = E(\left[ (|A| - |\Omega|) + \frac{1}{2}(|a| - |\omega|) \right]^2) \,, \tag{146}$$

$$= E([|A| - |\Omega|]^2) + E([|A| - |\Omega|][|a| - |\omega|]) + \frac{1}{4} E([|a| - |\omega|]^2) \,, \tag{147}$$

$$= E(|A|^2) - |\Omega|^2 + E(|A||a|) - |\Omega||\omega| + \frac{1}{4} \left\{ E(|a|^2) - |\omega|^2 \right\} \,, \tag{148}$$

$$= \frac{1}{n} \left\{ |\Omega|(1 - |\Omega|) - |\Omega||\omega| + \frac{1}{4}|\omega|(1 - |\omega|) \right\} \,. \tag{149}$$

A little algebra will show (149) is equal to

$$E([\ell - L]^2) = \frac{1}{n} \left\{ L \cdot (1 - L) - \frac{1}{4}|\omega| \right\} \leq \frac{L \cdot (1 - L)}{n} \,. \tag{150}$$

which is the other thing we wanted to prove.

## *Infinite and continuous spaces.*

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

$$T_4(X) : E(L(x)) = \frac{1}{2} \,, \tag{151}$$

$$T_5(X) : \left| E(f(L)) - \int_0^1 f(L)\, dL \right| \leq \frac{\max_L |f''(L)| \cdot \sup_x |\omega(x)|^2}{24} \,, \tag{152}$$

$$T_6(X) : 0 \leq \frac{1}{p+1} - E(L^p) \leq \frac{p \cdot (p-1) \cdot \sup_x |\omega(x)|^2}{24} \,. \tag{153}$$

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.

# *Index*