COUNTEREXAMPLES IN PROBABILITY AND STATISTICS

John Kermond

Haileybury College Senior School (Keysborough Campus)

A counterexample is an example or result that is counter to intuition or commonly held beliefs. It is used to disprove an incorrect statement or conjecture. In this paper some incorrect beliefs commonly held by secondary school students in probability and statistics are presented. Simple counterexamples that disprove these beliefs are provided and discussed in detail.

Introduction

A counterexample is an example or result that is counter to intuition or commonly held beliefs. It can be a powerful way of disproving an incorrect statement or conjecture.

In this paper some incorrect beliefs commonly held by secondary school students in probability and statistics are presented. Simple counterexamples that disprove these beliefs are provided and discussed in detail. Counterexamples that are more technical and apply to more advanced areas of probability and statistics can be found in Romano and Siegel (1986).

Misconception 1: A continuous probability distribution does not have a mode

By definition, the mode of a continuous probability distribution is the value at which its probability density function (pdf) attains its maximum value (Romano and Siegel 1986). Counterexamples readily follow from this definition. For example, the mode of a normal distribution is its mean. "The mode may not be uniquely defined if the maximum density is achieved over an interval, such as the mode of a uniform distribution." (HREF1).

There is no consensus on whether the uniform distribution is weakly unimodal (where any element of its support can be taken as a mode (see for example HREF2)) or has no mode (see for example Attwood *et al* 2000), suggesting perhaps the need for a more rigorous definition.

Misconception 2: Every probability distribution has a mean and a variance

Counterexample 1

A continuous random variable X has a *Cauchy distribution* if it has a pdf given by

$$f(x) = \frac{\alpha}{\pi}\frac{1}{(x-x_0)^2+\alpha^2}, \qquad -\infty < x < +\infty$$

where $x_0 \in R$ and $\alpha > 0$ are constants. The Cauchy distribution has a central peak (the parameter x_0 specifies its centre) and is symmetric (the parameter α specifies its width). It can be shown that the median and the mode are both equal to x_0. It can also be shown that the mean and variance do not exist. For example, consider the continuous random variable X that has a standard Cauchy distribution. Then $f(x) = \dfrac{1}{\pi(x^2+1)}$ and so

$E(X) = \dfrac{1}{\pi}\displaystyle\int_{-\infty}^{+\infty}\frac{x}{x^2+1}\,dx$. It is tempting to assign the value of zero to this *improper integral* because the integrand is an odd function:

$$\int_{-\infty}^{+\infty}\frac{x}{x^2+1}\,dx = \lim_{a\to+\infty}\int_{-a}^{a}\frac{x}{x^2+1}\,dx = \lim_{a\to+\infty}(0) = 0$$

(this is the Cauchy principle value). However, the integral could also be taken as:

$$\int_{-\infty}^{+\infty}\frac{x}{x^2+1}\,dx = \lim_{a\to+\infty}\int_{-2a}^{a}\frac{x}{x^2+1}\,dx = \frac{1}{2}\lim_{a\to+\infty}\left[\log_e\left(x^2+1\right)\right]_{-2a}^{a}$$

$$= \frac{1}{2}\lim_{a\to+\infty}\left(\log_e\frac{a^2+1}{4a^2+1}\right) = \frac{1}{2}\log_e\lim_{a\to+\infty}\left(\frac{a^2+1}{4a^2+1}\right) = -\log_e(2)$$

and so there is a contradiction. In fact, for $\displaystyle\int_{-\infty}^{+\infty}\frac{x}{x^2+1}\,dx$ to exist, both the improper integrals $\displaystyle\int_{-\infty}^{b}\frac{x}{x^2+1}\,dx$ and $\displaystyle\int_{b}^{+\infty}\frac{x}{x^2+1}\,dx$ need to be finite for all $-\infty < b < +\infty$ (Apostol

1981 p277) and this is not the case. For example:

$$\int_0^{+\infty} \frac{x}{x^2+1}\,dx = \lim_{a\to+\infty}\int_0^{a} \frac{x}{x^2+1}\,dx = \frac{1}{2}\lim_{a\to+\infty}\Big[\log_e\left(x^2+1\right)\Big]_0^a = \frac{1}{2}\lim_{a\to+\infty}\log_e\left(a^2+1\right) = +\infty$$

$$\int_{-\infty}^{0} \frac{x}{x^2+1}\,dx = \lim_{a\to-\infty}\int_a^{0} \frac{x}{x^2+1}\,dx = \frac{1}{2}\lim_{a\to-\infty}\Big[\log_e\left(x^2+1\right)\Big]_a^0 = -\frac{1}{2}\lim_{a\to-\infty}\log_e\left(a^2+1\right) = -\infty$$

Therefore the mean does not exist. It follows that the variance does not exist (since variance is defined with respect to the mean).

The fact that the Cauchy distribution has no mean can also be understood by noting that the tails of the Cauchy pdf approach the axis very slowly (it is a 'fat-tailed' distribution). This indicates higher probabilities for extremely large or small values and as a consequence the mean does not exist. It is interesting to note that if the standard normal distribution is drawn to scale on a sheet of paper so that its ordinate at $z = 6$ is 1mm high, then the corresponding standard Cauchy ordinate would be nearly 1·4 km high.

An implication of the mean not existing is that if a random sample $x_1, x_2, \ldots x_n$ is taken from the Cauchy distribution, then the limit of the average $\bar{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$ as the sample size increases does not exist (see HREF3). Technically this is a corollary of the fact that the sum of n independent Cauchy random variables is also a Cauchy random variable (Springer 1979).

Applications of the Cauchy distribution include ".... analysis of earthquake fault mechanisms, [explaining] the dispersion in the regional orientation of fault ruptures." (Woo 1999, p. 88). Other applications can be found in Krishnamoorthy (2006).

Counterexample 2

A continuous random variable X has a ***generalised Pareto distribution*** if it has a pdf given by

$$f(x) = \begin{cases} \dfrac{1}{\sigma}\dfrac{1}{(\xi z+1)^{1+\frac{1}{\xi}}}, & x \ge k \text{ when } \xi \ge 0 \text{ and } k \le x \le k - \frac{\sigma}{\xi} \text{ when } \xi < 0 \\ 0 & \text{otherwise} \end{cases}$$

where $z = \dfrac{x-k}{\sigma}$ and $\sigma > 0$, $k \in R$ and $\xi \in R$ are constants that specify scale, location and shape respectively. Note that $\lim\limits_{\xi\to 0} f(x) = \dfrac{1}{\sigma}e^{-(x-k)/\sigma}$:

$$\lim_{\xi\to 0}\frac{1}{(\xi z+1)^{1+\frac{1}{\xi}}}=\lim_{\xi\to 0}\frac{1}{(\xi z+1)}\cdot\lim_{\xi\to 0}(\xi z+1)^{-\frac{1}{\xi}}=\lim_{t\to 0}\left(1-\frac{z}{t}\right)^{t}=e^{-z}$$

It can be shown that the mean does not exist for $\xi\geq 1$. For example, consider the continuous random variable X that has a generalized Pareto distribution with $k = 0$ and $\xi = 1$. Then

$$f(x)=\begin{cases}\dfrac{\sigma}{(x+\sigma)^2}, & x\geq 0\\ 0 & \text{otherwise}\end{cases}$$

$$E(X)=\sigma\int_0^{+\infty}\frac{x}{(x+\sigma)^2}\,dx=\sigma\lim_{a\to+\infty}\int_\sigma^a\frac{u-\sigma}{u^2}\,du=\sigma\lim_{a\to+\infty}\left[\log_e|u|+\frac{\sigma}{u}\right]_\sigma^a=+\infty$$

and so the mean does not exist.

Counterexample 3

Consider the discrete random variable X that has a probability mass function (pmf) given by $p(x)=\dfrac{3}{\pi^2x^2}$ where $x=\pm 1,\ \pm 2,\ \pm 3\ \ldots$. Note that this is a pmf since $0\leq p(x)\leq 1$ and $\dfrac{3}{\pi^2}\sum_{x=-\infty}^{+\infty}\dfrac{1}{x^2}=\dfrac{3}{\pi^2}\left(\sum_{x=-\infty}^{-1}\dfrac{1}{x^2}+\sum_{x=1}^{+\infty}\dfrac{1}{x^2}\right)=\dfrac{3}{\pi^2}\left(\dfrac{\pi^2}{6}+\dfrac{\pi^2}{6}\right)=1$.

$$E(X)=\frac{3}{\pi^2}\sum_{x=-\infty}^{+\infty}\frac{1}{x}=\frac{3}{\pi^2}\left(\sum_{x=-\infty}^{-1}\frac{1}{x}+\sum_{x=1}^{+\infty}\frac{1}{x}\right)=\frac{3}{\pi^2}\left(\sum_{x=-\infty}^{-1}\frac{1}{x}+\sum_{x=1}^{+\infty}\frac{1}{x}\right)=\frac{3}{\pi^2}\left(\sum_{x=1}^{+\infty}\frac{1}{x}-\sum_{x=1}^{+\infty}\frac{1}{x}\right)=+\infty-\infty.$$

Therefore the mean does not exist.

The above distribution is an example of the ***Zipf distribution***. The Zipf distribution is commonly used in linguistics, insurance and the modeling of rare events (Krishnamoorthy 2006).

Misconception 3: A probability distribution that has a mean will have a (finite) variance

Counterexample

A continuous random variable X has a ***Pareto distribution*** if it has a pdf given by

$$f(x)=\begin{cases}\dfrac{\alpha k^\alpha}{x^{\alpha+1}}, & x\geq k\\ 0 & x<k\end{cases}$$

where $\alpha > 0$ and $k > 0$ are constants. This is a special case of the generalised Pareto distribution where $\xi = \frac{1}{\alpha} > 0$ and $k = \frac{\sigma}{\xi}$.

It can be shown that the mean exists for $\alpha > 1$ and that the variance exists for $\alpha > 2$. Therefore the distribution has a mean but not a variance when $1 < \alpha \le 2$. For example, consider the continuous random variable X that has a Pareto distribution with $\alpha = 2$ and $k = 1$:

$$E(X) = 2\int_1^{+\infty} \frac{1}{x^2}\,dx = 2\lim_{a\to+\infty}\int_1^{a} \frac{1}{x^2}\,dx = -2\lim_{a\to+\infty}\left[\frac{1}{x}\right]_1^a = -2\lim_{a\to+\infty}\left(\frac{1}{a}-1\right) = 2 .$$

$$E\left(X^2\right) = 2\int_1^{+\infty} \frac{1}{x}\,dx = 2\lim_{a\to+\infty}\int_1^{a} \frac{1}{x}\,dx = 2\lim_{a\to+\infty}\left[\log_e |x|\right]_1^a = 2\lim_{a\to+\infty}\left[\log_e |a|\right] = +\infty .$$

Therefore $Var(X) = E\left(X^2\right) - [E(X)]^2$ is infinite.

An implication of infinite variance is that the random variable associates with high variability in its distribution. De Vany and Walls (1999) found the Pareto distribution to be an excellent model for the box-office revenue of movies in the USA. For all movies whose box-office revenue was greater than or equal to \$40 million ($k = 40$ million) they calculated $\alpha = 1\cdot 91$ (implying a finite mean and infinite variance). For movies with and without stars and $k = 40$ million they calculated $\alpha = 1\cdot 72$ and $\alpha = 2\cdot 26$ respectively. The mean of box-office revenue can be forecast "since it exists and is finite, but the confidence interval of the forecast is without bounds. ... The sample variance is unstable and [very much] less than the theoretical value ... [For] movies without stars [the] variance is 122 million times as large as the expectation." (HREF4, pp. 3, 8, 20). Revenue forecasts therefore have zero precision because in each case the size of the variance completely overwhelms the value of the forecast.

The Pareto distribution is used primarily in the business and economics fields (it was originally used to describe the distribution of wealth among individuals) but also has actuarial, scientific and geophysical applications (Krishnamoorthy 2006).

Counterexample 2

Consider the discrete random variable X that has a pmf given by $p(x) = \dfrac{1}{\varsigma(3)x^3}$ where $x = 1,\ 2,\ 3\ \ldots$ (this is an example of the ***Zipf distribution***). $\varsigma(3) = \sum_{k=1}^{+\infty} \frac{1}{k^3}$ is known as

Apéry's constant (Finch 2003 p. 40).

$$E(X) = \frac{1}{\varsigma(3)} \sum_{x=1}^{+\infty} \frac{1}{x^2} = \frac{\pi^2}{6\varsigma(3)} .$$

$$E\left(X^2\right) = \frac{1}{\varsigma(3)} \sum_{x=1}^{+\infty} \frac{1}{x} = +\infty .$$

Therefore $Var(X) = E\left(X^2\right) - [E(X)]^2$ is infinite.

Misconception 4: Increasing the size of a data set will always reduce the uncertainty in the estimate of a parameter

Let S be an isotropic source emitting particles in the plane and let D be a line at unit distance from S. The angle of emission $\theta = \cot^{-1}(x)$ is a random variable (see Figure 1). A scientist trying to estimate the x-position of S relative to an origin O measures the x-position of a certain number of particle impacts on D and considers using their average value as an estimate of the x-position of S.

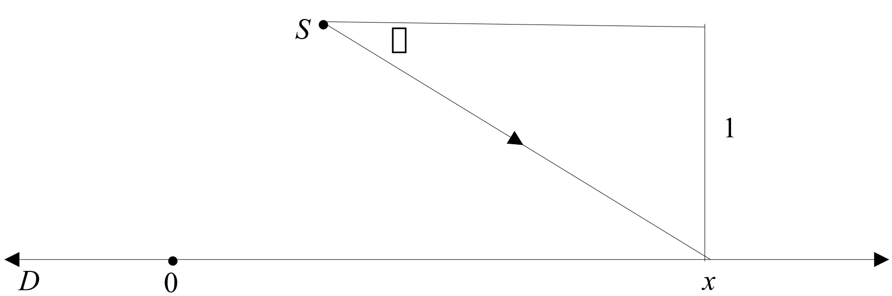

Figure 1. An isotropic source S *emitting particles in the plane.* D *is a line at unit distance from* S. *The angle of emission* $\theta = \cot^{-1}(x)$ *is a random variable.*

But if θ is uniformly distributed between 0 and π so that it has a pdf given by

$$f(\theta) = \begin{cases} \frac{1}{\pi}, & 0 < \theta < \pi \\ 0 & \text{otherwise} \end{cases} ,$$

then $X = \cot(\theta)$ is a random variable that has a standard Cauchy distribution.

Proof: Cumulative distribution function of X:

$$G(x) = \Pr(X < x), \; -\infty < x < +\infty,$$

$$= \Pr(\cot(\theta) < x) = \Pr(\theta > \cot^{-1}(x)) = 1 - \Pr(\theta < \cot^{-1}(x))$$

$$= 1 - \int_0^{\cot^{-1}(x)} \frac{1}{\pi}\, d\theta = \left[\frac{\theta}{\pi}\right]_0^{\cot^{-1}(x)} = 1 - \frac{1}{\pi}\cot^{-1}(x).$$

Probability distribution function of X:

$$\boldsymbol{g}(\boldsymbol{x}) = \frac{\boldsymbol{dG}}{\boldsymbol{dx}} = -\frac{1}{\pi}\frac{\boldsymbol{d}}{\boldsymbol{dx}}\left[\cot^{-1}(\boldsymbol{x})\right] = \frac{1}{\pi(1+\boldsymbol{x}^2)}, \; -\infty < \boldsymbol{x} < +\infty.$$

Therefore the average $\bar{\boldsymbol{x}} = \dfrac{\boldsymbol{x}_1 + \boldsymbol{x}_2 + \ldots + \boldsymbol{x}_n}{\boldsymbol{n}}$ will not approach a limiting value as the sample size increases. This means that the size of the sample will make no difference as to the uncertainty about the x-position of S.

Misconception 5: Pairwise independence of a set of events implies independence of the events

Counterexample

Two events A_1 and A_2 are independent if and only if $\Pr(A_1 \cap A_2) = \Pr(A_1)\cdot\Pr(A_2)$. More generally, the events A_1, A_2, …., A_n are independent if and only if $\Pr(\boldsymbol{A}_1 \cap \boldsymbol{A}_2 \cap \ldots \cap \boldsymbol{A}_n) = \Pr(\boldsymbol{A}_1)\cdot\Pr(\boldsymbol{A}_2)\cdot\ldots\Pr(\boldsymbol{A}_n)$.

Consider now the tossing of a coin twice and let the following events be defined:

A = Heads on the first toss, B = Heads on the second toss, C = One head and one tail (in either order) is tossed.

Then from the sample space it is clear that:

$\Pr(A) = \Pr(B) = \Pr(C) = \frac{1}{2}$

$\Pr(A \cap B) = \Pr(B \cap C) = \Pr(C \cap A) = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2}$.

"It follows that any two of the events are pairwise independent. However, since the three events cannot all occur simultaneously, we have …." (Romano and Siegel 1986): $\Pr(A \cap B \cap C) = 0 \neq \frac{1}{8} = \Pr(A)\cdot\Pr(B)\cdot\Pr(C)$. Therefore the three events A, B and C are not independent.

Misconception 6: Something true of each subset of a population will be true of the population as a whole

Counterexample: Simpson's Paradox

"Simpson's paradox refers to the reversal of the direction of a comparison or an association when data from several groups are combined to form a single group." (Moore and McCabe 1993, p. 225). It is often encountered in sport, social science and medical science data and occurs when a causal relationship based on the association between two variables is inferred but the effect of a relevant lurking variable is overlooked. While Simpson's paradox often refers to the analysis of count tables, it can also occur with continuous data. For example, separated regression lines fitted through two sets of data may show a positive trend, while a single regression line fitted through all the data together may show a negative trend (see for example Rücker G. and Schumacher M. (2008), HREF5).

Example

Consider a medical study comparing the success rates of two treatments for kidney stones (Charig *et al* 1986, Julious and Mullee 1994). The overall success rates and numbers of treatments for each treatment is shown in table 1. This seems to show that treatment 2 is more effective than treatment 1. However, if data about kidney size is included (see Table 2) the conclusion about the effectiveness of each treatment is reversed. Treatment 1 is seen to be more effective for both small and large stones.

	Treatment 1	Treatment 2
Successful	273 (78%)	289 (83%)
Unsuccessful	77	61
Total	350	350

Table 1. Overall success rates and numbers of treatments for two treatments for kidney stones.

	Small Stones		Large Stones	
	Treatment 1	Treatment 2	Treatment 1	Treatment 2
Successful	81 (93%)	234 (87%)	192 (73%)	55 (69%)
Unsuccessful	6	36	71	25
Total	87	270	263	80

Table 2. Success rates and numbers of treatments for two treatments for kidney stones by kidney stone size.

The lurking variable is the size of the kidney stone. Doctors tended to give the severe cases (that is, large kidney stones) the superior treatment (treatment 1) and the milder cases (that is, small stones) the inferior treatment (treatment 2). Since the severity of the case can influence the success of a treatment, this difference in kidney stone size reduces the success rate of treatment 1 despite its superior effectiveness on each size of kidney stone.

Mathematical resolution of the paradox

It is possible to have $\Pr(A \mid B) < \Pr(A \mid B')$ and at the same time have

$\Pr(A \mid B \cap C) \geq \Pr(A \mid B' \cap C)$

$\Pr(A \mid B \cap C') \geq \Pr(A \mid B' \cap C')$

where C represents the lurking variable. People tend to reason intuitively that this is impossible because

$$\Pr(A \mid B) \text{ is an average of } \Pr(A \mid B \cap C) \text{ and } \Pr(A \mid B \cap C')$$
$$\Pr(A \mid B') \text{ is an average of } \Pr(A \mid B' \cap C) \text{ and } \Pr(A \mid B' \cap C').$$

Although this is true, the reasoning fails because these two averages have **different weightings**:

$$\Pr(A \mid B) = \underline{\Pr(C \mid B)} \cdot \Pr(A \mid B \cap C) + \underline{\Pr(C' \mid B)} \cdot \Pr(A \mid B \cap C')$$
$$\Pr(A \mid B') = \underline{\Pr(C \mid B')} \cdot \Pr(A \mid B' \cap C) + \underline{\Pr(C' \mid B')} \cdot \Pr(A \mid B' \cap C')$$

The reasoning is only correct when *B* and *C* are independent and therefore the weightings are the same for each average. It would then follow from $\Pr(A \mid B \cap C) \geq \Pr(A \mid B' \cap C)$ and $\Pr(A \mid B \cap C') \geq \Pr(A \mid B' \cap C')$ that $\Pr(A \mid B) > \Pr(A \mid B')$ and there is no paradox.

In the previous example, let the following propositions be defined:

A = The treatment is successful, B = Treatment 1 is used, C = The kidney stone is small.

- $\Pr(A \mid B) = 0.78$, $\Pr(A \mid B') = 0.83$ and so $\Pr(A \mid B) < \Pr(A \mid B')$.
- $\Pr(C \mid B) = \frac{87}{87+263} = \frac{87}{350}$ and $\Pr(C' \mid B) = \frac{263}{263+87} = \frac{263}{350}$. Therefore:
  $\Pr(A \mid B) = \frac{87}{350} \cdot \Pr(A \mid B \cap C) + \frac{263}{350} \cdot \Pr(A \mid B \cap C')$.
- $\Pr(C \mid B') = \frac{270}{270+80} = \frac{270}{350}$ and $\Pr(C' \mid B') = \frac{80}{80+270} = \frac{80}{350}$. Therefore:
  $\Pr(A \mid B') = \frac{270}{350} \cdot \Pr(A \mid B' \cap C) + \frac{80}{350} \cdot \Pr(A \mid B' \cap C')$.
- $\Pr(A \mid B \cap C) = \frac{81}{87}$ and $\Pr(A \mid B' \cap C) = \frac{234}{270}$. Therefore:
  $\Pr(A \mid B \cap C) \geq \Pr(A \mid B' \cap C)$.
- $\Pr(A \mid B \cap C') = \frac{192}{263}$ and $\Pr(A \mid B' \cap C') = \frac{55}{80}$. Therefore:
  $\Pr(A \mid B \cap C') \geq \Pr(A \mid B' \cap C')$.

There would be no paradox if *B* and *C* were independent, that is, if the proportion of cases receiving treatment 1 was the same for small stones and large stones. However, for the given data the proportion of small stone and large stone cases receiving treatment 1 is $\frac{87}{357}$ and $\frac{263}{343}$ respectively.

Misconception 7: Pr(*A* | *B*) = Pr(*B* | *A*)

The Prosecutors Fallacy

The *Prosecutors Fallacy* is a mis-statement of probability as a result of a misunderstanding of conditional probability and often occurs in legal arguments. For example (Matthews 1997), when the prosecution interprets data about DNA and blood group evidence in the wrong way. A so-called match probability of 1 in 40 million represents the chances of getting so good a match assuming that the defendant is innocent: Pr(match | innocent). However, what the jury is trying to decide is the probability of innocence given the DNA information: Pr(innocent | match). The jury is misled by the prosecution into thinking that the 1 in 40 million match probability represents the chances of the defendant being innocent.

Let the following propositions be defined:

A = The defendant is innocent, I = All prior information (defendant had an alibi, defendant not identified in line-up etc.), D = DNA evidence matches the defendant.

From Bayes' Theorem:

$$\Pr(A \mid D \cap I) = \frac{\Pr(A \cap D \mid I)}{\Pr(D \mid I)} = \frac{\Pr(A \mid I) \cdot \Pr(D \mid A \cap I)}{\Pr(D \mid I)}$$

$$= \frac{\Pr(A \mid I) \cdot \Pr(D \mid A \cap I)}{\Pr(D \mid A \cap I) \cdot \Pr(A \mid I) + \Pr(D \mid A' \cap I) \cdot \Pr(A' \mid I)}.$$

It's given that $\Pr(D \mid A \cap I) = \frac{1}{40{,}000{,}000}$. Assume that $\Pr(D \mid A' \cap I) = 1$. Then:

$$\Pr(A \mid D \cap I) = \frac{\Pr(A \mid I) \cdot \frac{1}{40{,}000{,}000}}{\frac{1}{40{,}000{,}000} \cdot \Pr(A \mid I) + \Pr(A' \mid I)} = \frac{1}{1 + 40{,}000{,}000 \cdot \frac{\Pr(A' \mid I)}{\Pr(A \mid I)}}.$$

If for example $\Pr(A' \mid I) = \frac{1}{3{,}000{,}000}$ so that $\Pr(A \mid I) = \frac{2{,}999{,}999}{3{,}000{,}000}$ then the probability of innocence given all the evidence is $\Pr(A \mid D \cap I) \approx \frac{7}{100}$ and there is clearly reasonable doubt. The prosecutor, however, would have the jury believe that the probability of innocence is $\Pr(D \mid A \cap I) = \frac{1}{40{,}000{,}000}$.

Counterexample

If a card is selected from a standard deck of cards, Pr(King | picture card) ≠ Pr(picture card | King).

References

Apostol T. M. (1981). *Mathematical Analysis.* Reading, Massachusetts: Addison-Wesley Publishing Company.

Attwood G. *et al* (2000). *Statistics 2 (Heinemann Modular Mathematics for Edexcel AS & A Level).* Oxford: Heinemann Educational Publishers.

Blyth, C. R. (1972). On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association, 67* (338), 364-366.

Charig, C. R. *et al* (1986). Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. *British Medical Journal (Clinical Research Edition), 292,* 879-882.

De Vany, A. and W. D. Walls (1999). Uncertainty in the movie industry: Does star power reduce the terror of the box office? *Journal of Cultural Economics, 23* (4), 285-318.

Finch, S. R. (2003). *Mathematical Constants.* Cambridge University Press.

Julious, S and Mullee, M. (1994). *Confounding and Simpson's paradox.* British Medical Journal, 309, 1480-1481.

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions With Applications.* Boca Raton: Chapman and Hall/CRC.

Matthews, R. (1997). *Tipping the scales of justice.* New Scientist, 13 December 1997, 18-19.

Moore, D. S. and McCabe, G. P. (1993). *Introduction to the Practice of Statistics* (2nd edition). New York: W. H. Freeman & Company.

Romano, J. P. and Siegel, A. F. (1986). *Counterexamples in Probability and Statistics.* Boca Raton: Chapman and Hall/CRC.

Rücker, G. and Schumacher, M. (2008). Simpson's paradox visualized: The example of the Rosiglitazone meta-analysis. *BMC Medical Research Methodology,* 8, 34-42.

Woo, G. (1999). *The Mathematics of Natural Catastrophes.* London: Imperial College Press.

Springer, M. D. (1979). *The Algebra of Random Variables*. New York: John Wiley and Sons.

HREF1: **http://www.btinternet.com/~se16/hgb/mode.pdf**. Last visited 08/08/2009.

HREF2: **http://en.wikipedia.org/wiki/Uniform_distribution_(continuous)**. Last visited 08/08/2009.

HREF3: **http//:www.johndcook.com/Cauchy_estimation.html.** Last visited 08/08/2009.

HREF4: **http://pages.stern.nyu.edu/~wgreene/entertainmentandmedia/Devany& Walls.pdf**. Last visited 08/08/2009.

HREF5: **http://ukpmc.ac.uk/articlerender.cgi?artid=1561994**. Last visited 08/08/2009.