Noisy Reasoning: a Model of Probability Estimation and Inferential Judgment

Fintan Costello,[1*] Paul Watts[2]

[1]School of Computer Science and Informatics
University College Dublin, Belfield, Dublin 4, Ireland
[2]Department of Mathematical Physics
National University of Ireland, Maynooth, Co Kildare, Ireland

[*]To whom correspondence should be addressed; E-mail: fintan.costello@ucd.ie.

Abstract

We describe a computational model of two central aspects of people's probabilistic reasoning: descriptive probability estimation and inferential probability judgment. This model assumes that people's reasoning follows standard frequentist probability theory, but is subject to random noise. This random noise has a regressive effect in probability estimation, moving probability estimates away from normative probabilities and towards the center of the probability scale. This regressive effect explains various reliable and systematic biases seen in people's probability estimation. This random noise has an anti-regressive effect in inferential judgment, however. This model predicts that these contrary effects will tend to cancel out in tasks that involve both descriptive probability estimation and inferential probability judgment, leading to unbiased responses in those tasks. We test this model by applying it to one such task, described by Gallistel et al. (2014). Participants' median responses in this task were unbiased, agreeing with normative probability theory over the full range of responses. Our model captures the pattern of unbiased responses in this task, while simultaneously explaining systematic biases away from normatively correct probabilities seen in other tasks.

We live in a world of nonstationary stochastic processes, where events occur with some associated probability, and this probability itself changes unpredictably over time. To make successful predictions about event occurrence in such a world we must use two distinct types of probabilistic reasoning: descriptive probability estimation (given the events we have seen recently, what is the current underlying probability of A?) and inferential probability judgment (given our current estimate for the probability of A, is the current sample of events consistent with that probability? Or should we infer that the underlying probability of A has changed?). Our aim in this paper is to present a computational model of these two interacting components of probabilistic reasoning.

One revealing aspect of human probabilistic reasoning is the reliable occurrence of a number of systematic biases; biases such as conservatism (Erev et al., 1994), subadditivity (Tversky and Koehler, 1994) and the conjunction fallacy (Tversky and Kahneman, 1983). The model we present here was originally developed to explain these biases in terms of the regressive effect of random noise in reasoning (see Costello and Watts, 2014). Here we extend this model to inferential probability judgment, and show that this model explains patterns of bias seen in such judgment. This model predicts that, in situations that involve both forms of reasoning, these regressive effects will tend to cancel out, leaving subjective probability estimates that tend to agree with the normatively correct values with no systematic bias. Such agreement is seen in recent studies of probability estimation for nonstationary stochastic processes by Gallistel et al. (2014). We demonstrate the model by applying it to Gallistel et al.'s study in detail.

The probability theory plus noise model

Our model assumes that people's probability judgments are produced by a mechanism that is fundamentally rational, but is perturbed in various ways by purely random noise or error, which causes systematic regressive effects. We take $P(A)$ to represent the 'true' probability of event A (that is, the proportion of items in memory that represent A). We take $p_*(A)$ to represent an individual estimate of the probability of event A, and take $\langle p_*(A)\rangle$ to represent the expectation value or mean of these estimates for A: this is the value we would expect to get if we averaged an infinite number of individual estimates for $p_*(A)$. In standard probability theory, the probability of some event A is estimated by drawing a random sample of events, counting the number of those events that are instances of A, and dividing by the sample size. The expected value of these estimates is $P(A)$, the probability of A. We assume that people estimate the probability of some event A in exactly this way: randomly sampling events from memory, counting the number of instances of A, and dividing by the sample size.

If this counting process was error-free, people's estimates would have an expected value of $P(A)$. Human memory, however, is subject to various forms of random error or noise. To reflect this we assume events have some chance $d < 0.5$ of randomly being counted incorrectly: there is a chance d that a $\neg A$ (*not* A) event will be incorrectly counted as A, and the same chance d that an A event will be incorrectly counted as $\neg A$. Given this form of noise, a randomly sampled event will be counted as A if the event truly is A and is counted correctly (with a probability $(1-d)P(A)$, since $P(A)$ events are truly A and events have a $1-d$ chance of being counted correctly), or if the event is truly $\neg A$ and is counted incorrectly as A (with a probability $(1-P(A))d$, since $1-P(A)$ events are truly $\neg A$, and events have a d chance of being counted incorrectly). Summing the probabilities of these two mutually exclusive situations, we get an expected value for a noisy probability estimate of

$$\langle p_*(A)\rangle = (1-2d)P(A)+d \qquad (1)$$

with individual estimates varying independently around this expected value. This average is systematically biased away from the 'true' probability $P(A)$, such that estimates will tend to be greater than $P(A)$ when $P(A) < 0.5$, and will tend to be less than $P(A)$ when $P(A) > 0.5$: a pattern of systematic regression towards 0.5, the center of the probability scale.

Regression, in this model, explains a number of observed patterns of bias in people's probability estimates, such as conservatism, subadditivity, and the conjunction fallacy (see Costello and Watts, 2016a, 2014). This model also makes a number of novel predictions about patterns of bias and agreement with probability theory for various probabilistic expressions; for example, this model predicts that

$$p_*(A) + p_*(B) - p_*(A \wedge B) - p_*(A \vee B) = 0$$

will hold, on average, in people's probability estimates for any events A and B (because in this expression the regressive effects of noise on individual probability estimates $p_*(A)$, $p_*(B)$ $p_*(A \wedge B)$ and $p_*(A \vee B)$ will tend to cancel out). These predictions are strongly supported by experimental results (see Costello and Watts, 2014, 2016b).

Inferential probability judgment

Equation 1 describes the expected value for a probability estimate in one type of probabilistic reasoning task: one where the reasoner sees a sample containing some instances for the event of interest, A, and produces an estimate of the underlying probability $P(A)$. This type of task involves the estimation of a descriptive probability: a probability that summarises the observed sample. We now consider a probabilistic reasoning task where the reasoner is given an explicit probability value p and a sample of n events containing x instances of event A , and judges whether the number of A's seen in the sample is consistent with the given probability. This type of task involves the estimation of an inferential probability $P(x,n|P(A) = p)$: the probability of seeing x A's in a sample of n items, given that $P(A) = p$. Frequentist probability theory provides a normative mechanism for estimating such inferential probabilities: to estimate $P(x,n|P(A) = p)$, draw a series of random samples, each of size n, from a population where $P(A) = p$ and count the proportion of samples that contain exactly x instances of A. This proportion gives an estimate of the probability of the observed sample occurring in a population with $P(A) = p$: the lower this estimate, the less likely it is that the observed sample came from such a population. The expected value of this estimate is given by the binomial probability function

$$P(x,n|p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{2}$$

In our model we assume that people estimate inferential probabilities just as in frequentist probability theory: by drawing a series of random samples of size n from a (simulated) population where $P(A) = p$, and counting the proportion of samples that contain exactly x instances of A. We assume that this counting process is subject to random error; that the count of occurrences of A in a sample is subject to random noise at a rate d (there is d chance that an instance of A in a given sample will be counted as $\neg A$, and d chance that an instance of $\neg A$ in a given sample will be counted as A). Given this random error, with $P(A) = p$ the chance of an instance in a sample being counted as A is equal to $(1-2d)p+d$ (from Equation 1), and so the expected value for this noisy estimate is given by the binomial probability

$$\langle p_*(x,n|p) \rangle = \binom{n}{x} ((1-2d)p+d)^x ((1-2d)(1-p)+d)^{n-x} \tag{3}$$

Note that the probabilities given in Equation 2 and Equation 3 are both binomially distributed with common terms x and n. If we take p_e to be our current estimate of the probablity of A in the population in question, this means that, for any given values of x and n, the associated noisy inferential probability $\langle p_*(x,n|p_e) \rangle$ is exactly equal to another normatively correct inferential probability $P(x,n|p)$ when

$$((1-2d)p_e+d)^x((1-2d)(1-p_e)+d)^{n-x} = p^x(1-p)^{n-x}$$

When $d \leq p \leq 1-d$, this equality holds for all values of n and x when

$$(1-2d)p_e + d = p$$

or equivalently when

$$p_e = \frac{p-d}{1-2d}$$

This expression is 'anti-regressive', giving values for p_e that are closer to the boundaries 0 and 1 than values of p: p_e is greater than p when $p > 0.5$, and less than p when $p < 0.5$.

Properties of the model

In this section we apply the above model to two sets of experimental results: on conservatism in inferential probability judgment, and on probability estimation in tasks that mix probability estimation and inferential judgment.

Conservatism in inferential judgment

Experimental studies typically investigate inferential probability estimation indirectly, using the related concept of relative probability. These studies involve describing two populations containing complementary proportions of two different types of event. Participants are told that a population has been picked at random, and are then shown a sample of events drawn from the selected population and asked to assess the probability that the sample came from one population rather than the other. Typically these populations are 'bookbags' containing poker chips, with one bag containing, for example, 70% red chips and 30% black (this is the 'red bag'), and the other bag containing the complementary proportions: 30% red chips and 70% black (this is the 'black bag'). Participants are told the distribution of chips in each bag. They are then shown a sequence of n chips and asked, after seeing each chip, to estimate the probability that the sample came from the red bag rather than the black bag, or vice versa (the relative probability of one bag over the other; see Peterson and Beach, 1967, for examples).

Having seen a sample of n events containing x red chips, the normatively correct relative probability that the sample came from the red bag rather than the black bag is given by

$$R(x,n,p) = \frac{P(x,n|p)}{P(x,n|p)+P(x,n|1-p)} = \frac{1}{1+\left[\frac{1-p}{p}\right]^x \left[\frac{p}{1-p}\right]^{n-x}} \tag{4}$$

(since the proportion of red chips is p in the red bag, and $1-p$ the black bag). As participants proceed through these tasks they give relative probability estimates that follow the direction required by normative probability theory, but with values of these estimates being 'conservative': less extreme than the normatively correct values. This means that if participants see $x > n/2$ red chips in their sample, they give estimates for the probability that the sample came from the red bag that are greater than 0.5 but less than the normatively correct value, while if participants see $x > n/2$ black chips in their sample, they give estimates for the probability that the sample came from the black bag that are greater than 0.5 but less than the normatively correct value. In applying our model to this task we assume, without loss of generality, that red chips are most frequent in the sample and take $x > n/2$ to be the number of red chips in the sample of n events that have been seen, and assume $p > 0.5$ to be the proportion of red chips in the red bag (the bag that participants associate with the sample).

The estimated relative probability, in our model, of a seeing a sample of size n with x red chips coming from the red bag rather than the black bag is given by

$$R_E(x,n,p) = \frac{p_*(x,n|p)}{p_*(x,n|p)+p_*(x,n|1-p)}$$

Note that, since by assumption $p > 0.5$ and $x > n/2$, from Equation 3 we see that $p_*(x,n|p) > p_*(x,n|1-p)$ will tend to hold (subject, of course, to random error: more specifically, the higher the values of x and p the more likely it is that this inequality will hold). This means that $R_E(x,n,p)$ will be greater than 0.5, and these noisy relative probability estimates will follow the direction required by normative probability theory, just as seen in experiments.

For $p > .5$ this function $R_E(x,n,p)$ will be concave for all $x > n/2$ (since as x increases from $n/2$ the probability that the sample came from the red bag increases while the probability that the sample came from the black bag simultaneously falls). Since from Jensen's Inequality we have $\langle f(x)\rangle \leq f(\langle x\rangle)$ for concave functions (the expected value of a concave function is less than that function of the expected value of its argument), we get

$$\left\langle \frac{p_*(x,n|p)}{p_*(x,n|p)+p_*(x,n|1-p)} \right\rangle \leq \frac{\langle p_*(x,n|p)\rangle}{\langle p_*(x,n|p)\rangle + \langle p_*(x,n|1-p)\rangle}$$

and so, rearranging and substituting, we get

$$\langle R_E(x,n,p)\rangle \leq \frac{1}{1+\left[\frac{(1-2d)(1-p)+d}{(1-2d)p+d}\right]^x \left[\frac{(1-2d)p+d}{(1-2d)(1-p)+d}\right]^{n-x}} \tag{5}$$

Comparing Equations 4 and 5 we see that $\langle R_E(x,n)\rangle < R(x,n,p)$ when

$$\left[\frac{1+d\left(\frac{1}{p}-2\right)}{1+d\left(\frac{1}{1-p}-2\right)}\right]^x < \left[\frac{1+d\left(\frac{1}{p}-2\right)}{1+d\left(\frac{1}{1-p}-2\right)}\right]^{n-x} \tag{6}$$

Since by assumption we have $p > 0.5$ and $x > n/2$ we see that the inequality in equation 6 always holds, and so $0.5 < \langle R_E(x,n,p)\rangle < R(x,n,p)$: estimated relative probability follows the direction required by probability theory, but is conservative, just as observed in people's relative probability judgments. In other words, even though the expected values for the individual inferential probability judgments $\langle p_*(x,n|p)\rangle$ and $\langle p_*(x,n|1-p)\rangle$ are each anti-regressive relative to their corresponding normative values in this model, when combined to produce an overall estimate of relative probability, this estimate is regressive and so reproduces the pattern of conservatism seen in inferential judgment.

Combined estimation and judgment tasks

We finally describe how this model applies to tasks that involve both descriptive and inferential probability estimation. We consider an iterative task that involves the repeated updating of an estimate for the hidden probablity parameter (which may itself randomly change), given a sample of events presented outcome by outcome. People's performance in such tasks were investigated in an experiment by Gallistel et al. (2014), where participants gave repeated estimates of the hidden parameter, p, of a stepwise non-stationary Bernoulli process that controlled the colour of a circle being drawn from a concealed box. On each trial participants clicked a button to draw a new circle from the box. After being drawn, the circle evaporated, and participants could update their estimate for the hidden probability p. Participants were told that the box would sometimes be silently replaced by another box with a different value of p. Participants could update their estimates by either clicking a "The box has changed!" button (and then picking a new probability estimate), or by adjusting their current probability estimate, or by making no change.

There were two main results from this experiment. First, people's probability estimates were characterised by rapid changes in the estimated value in response to changes in the underlying hidden probability, separated by periods of small adjustments in the estimate (see Figure 1, left side). The speed of detection of a change in the underlying probability p depended on the degree of change: large changes in the underlying probability were detected more rapidly than smaller changes. The median latency for detection of a change in probability estimate in response to a change in the underlying probability was around 12 events in Gallistel et al. (2014).

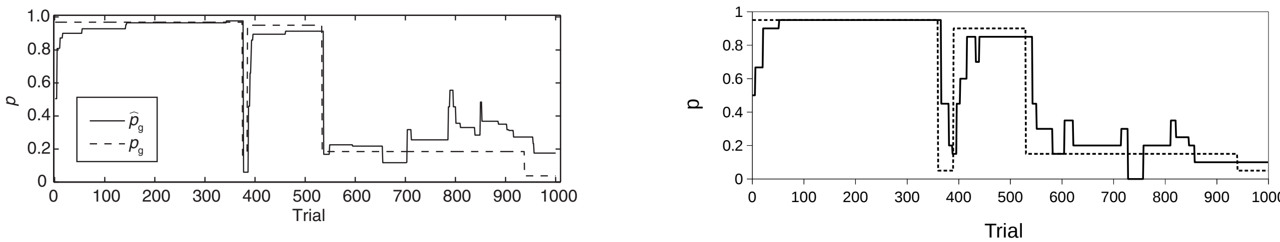

Figure 1. **(Left)** Trial-by-trial true probability (dashed line) and trial-by-trial probability estimate (solid line) for Subject 4, Session 8 in Gallistel et al.'s task (From Fig. 5 in Gallistel et al., 2014, page 102; p_g and $\hat{p}_g$ represent true and estimated probabilities respectively). **(Right)** Trial-by-trial probability estimates produced by our model for the same set of true probabilities. These graphs illustrate the step-hold pattern seen in Gallistel et al.'s task, and show that the model reproduces this pattern.

The second main result was that the relationship between the true probability p and participants estimated probability was essentially that of identity: the median trial-by-trial probability estimates closely tracked the true hidden probability with no systematic bias.

This pattern of agreement with the true probability arises, in our model, due to the cancellation of regressive effects in probability estimation against those in inferential judgment. Suppose we see a series of random samples drawn from a population with a parameter $p = P(A)$, and take p_e to represent our estimate of p (which we repeatedly update as outcomes are presented in the task). This estimate p_e will be subject to random noise, and so will have a regressive average value as in Equation 1. Individual estimates p_e will be adjusted (in a quasi-random walk) in response to inferential probability judgment of the chance of obtaining the currently-seen sequence of outcomes, given our current estimate. This inferential probability judgment will also be subject to random noise, and so will be anti-regressive. This estimate p_e will be least likely to be adjusted when it reaches a value maximally consistent with the average number of counted occurrences of A in the presented sample, and so will tend to fix at that value. Due to random noise, the average number of counted occurrences of A in a sample is equal to $[(1-2d)p+d]n$, and so p_e will fix at the value for which the inferential probability $\langle p_*([(1-2d)p+d]n, n|p_e)\rangle$ is maximised. Since from Equation 3 this inferential probability has a binomial distribution with probability $(1-2d)p_e+d$, it has its maximum value when

$$(1-2d)p_e + d = (1-2d)p + d$$

or equivalently, when $p_e = p$; when our estimate p_e for the underlying population probability equals the true value. In other words, even though descriptive probability estimates are regressive in this model (due to random noise), and inferential probability estimates are anti-regressive (also due to random noise), when these two types of probability judgment are combined these regressive and anti-regressive effects should on average cancel out, leaving estimates that on average agree with the hidden probability parameter p; just as seen in mixed estimation and inferential judgment tasks such as Gallistel et al.'s.

Computational simulation

We apply the model to Gallistel et al.'s continuous probability perception task by assuming that a continuous probability estimate p_e is assessed by counting the frequency of A in n just-observed events (subject to random noise). The parameter n here represents the size of short-term memory available to store just-seen events: we assume n is small, but beyond that make no assumptions about the value n (in our simulations, below, we chose n randomly for each simulated participant, uniformly in the range $5 \ldots 20$).

We take x to represent the number of occurrences of A in the n most recently observed events and take x_e to represent the noisy count of that number (the count of occurrences obtained with a chance d of randomly miscounting). The expected value of x_e equals $(1-2d)x+nd$, and so the immediately observed probability of A in that sample has the expected value

$$q = (1-2d)\frac{x}{n} + d \qquad (7)$$

On each event occurrence the model makes one of three choices, corresponding to the 3 choices available to participants in Gallistel et al.'s experiment. First, the model may reject the current value of p_e as inconsistent with the number of A's just observed, and update to a new estimate by setting $p_e = q$ (this choice corresponds to clicking "The box has changed!" in Gallistel et al.'s experiment). Second, the model may decide that the underlying distribution has *not* changed but that q is more consistent with the observed number of A's than p_e. In this case the model again updates to a new estimate by setting $p_e = q$: this choice corresponds to a small adjustment of the current probability estimate. Third, the model may decide not to modify p_e.

To decide whether the current estimate p_e needs to be rejected, the model considers the chance of seeing x_e occurrences of A in n samples where the probability of seeing A in those samples is actually p_e. If this chance is too low p_e is rejected. The model assesses this chance in a simple way: by generating 100 random samples (each of size n, with A occurring randomly with probability p_e) and counting the number of A's in each sample. This counting process is subject to random error, with some probability $d < 0.5$ that an occurrence of A will be counted as $\neg A$, or an occurrence of $\neg A$ will be counted as A. The proportion of these samples that contain exactly x_e occurrences of A represents an estimate of the inferential probability $P_E(x_e, n|p_e)$. If this inferential probability is less than some decision criterion T_1 the model concludes that p_e should be rejected because the underlying distribution has changed. The model then changes the new estimate to q.[1].

If the current estimate is not rejected, the model next considers making an estimate adjustment. To decide whether the current estimate p_e needs to be adjusted, the model considers the inferential probability $P_E(x_e, n|q)$: the chance of seeing x_e occurrences of A in n samples drawn from a population where $P(A) = q$. As above, the model assesses this chance by generating 100 random samples (each of size n, with A occurring randomly with probability q) and counting the number of A's in each sample (subject to a rate d of random error in counting). The proportion of these samples that contain exactly x_e occurrences of A represents an estimate of the inferential probability $P_E(x_e, n|q)$. If the difference between this inferential probability and the previous inferential probability is greater than some decision criterion T_2 (that is, if $P_E(x_e, n|q) - P_E(x_e, n|p_e) > T_2$) the model decides that q is a better estimate and changes to a new estimate by setting $p_e = q$. Otherwise the current estimate p_e is left unchanged.

Results

We wrote a computer program implementing this model and tested it by simulating Gallistel et al.'s experiment. On each run of this simulation the model was shown a consecutive sequence of 1000 randomly generated A or $\neg A$ events. After seeing each event, the model either rejected its current probability estimate and changed to the new estimate q; or adjusted its estimate to the new estimate q; or else left its estimate unchanged. Events were generated randomly, with a hidden probability p. The value of p itself changed randomly over the sequence of 1000 events, with the probability that p would change after a given event being set at a constant value of 0.005 (just as in Gallistel et al.'s experiment). The size and direction of a change in the hidden probability were determined by a random choice of the next value from a uniform distribution between 0 and 1, subject to the restriction that $p/(1-p)$, the resulting change in the odds, was no less than fourfold, just as in Gallistel et al. (2014).

To investigate the role of error in descriptive probability estimation and in inferential judgment, we designed the program so that we could set one error rate d for descriptive estimation, and another rate d_s for inferential judgment. We simulated Gallistel et al.'s experiment for 4 different pairs of values for these parameters: Sim A ($d = 0.0, d_s = 0.0$), Sim B ($d = 0.1, d_s = 0.0$), Sim C ($d = 0.0, d_s = 0.1$), and Sim D ($d = 0.1, d_s = 0.1$). We set the criterion parameters T_1 and T_2 to 0.01 and 0.1 respectively in all simulations, since initial tests suggested that these values produced a reasonable rate of adjustment in the model's probability estimates. Each simulation involved 500 'participants' (runs of the model), all with the same values for parameters d and d_s, and each with a value of n (the size of short-term memory) selected randomly from the range $5 \ldots 20$. Each 'participant' saw a different randomly generated sequence of 1000 events, produced according to a different randomly generated sequence of values of p (as in Gallistel et al., 2014).

Rapid detection of changes The median latency between a change in the hidden probability p and the recognition of that change by the model (via rejection of the current probability estimate) was 10 in simulations A and B, 13 in simulation C and 12 in simulation D. These values agree with the median latency of reported change detection of 12 seen in Gallistel et al. (2014).

High hit rates and low false alarm rates Gallistel et al. (2014) describe a method for computing hit rates and false-alarm rates in participant's responses in their experiment: they found that nine out of ten participants had hit rates in the range $0.77 \ldots 1$ and false-alarm rates in the range $0.004 \ldots 0.02$. We used the same method to compute hit rates and false alarm rates across all 'participants' in our simulations. Average hit rates were $0.87, 0.79, 0.81, 0.76$ and false-alarm rates were $0.006, 0.005, 0.005, 0.005$ in simulations A, B, C and D respectively. These agree with the rates seen by Gallistel et al. (2014).

Precision We assess the precision of the model's probability estimates by computing the RMSD between the model's estimate at a given event against the true probability p at that event. These RMSD's between estimated and true probabilities were $0.15, 0.17, 0.17, 0.17$ for simulations A, B, C and D respectively. These were consistent with the corresponding RMSD's for participants in Gallistel et al.'s experiment, which ranged between 0.15 and 0.21.

These three aspects of the model are illustrated in the right of Figure 1. This figure shows trial-by-trial probability estimates produced by the model for one run, with parameter values $d = 0.1, d_s = 0.1, n = 20$. Values of the true probability p were controlled match those in Gallistel et al.'s example. Individual event occurrences in this run, however, were random, and did not follow the precise sequence of event occurrences in Gallistel et al. (2014). This figure shows that the model produces the step-hold pattern seen in Gallistel et al.'s task, with large changes in the estimate when the hidden probability changes, and small adjustments, or no changes, otherwise.

Identity between true probability and median estimates Recall that the noisy frequentist model predicts that noise will have different effects in different probability judgment tasks: when estimating a probability from a sample (descriptive probability estimation), noise will produce regressive effects; when estimating the likelihood of a sample given a probabilty (inferential probability judgment), noise will produce anti-regressive effects; and in tasks that involve both

[1] Note that our decision to use 100 random samples when estimating inferential probabilities here is essentially arbitrary: this number was chosen to allow us to use values for the decision criteria T_1 and T_2 that correspond to standard significance level values such as 0.01 and 0.05. Versions of the simulation that make use of much smaller numbers of samples give essentially the same results as seen here.

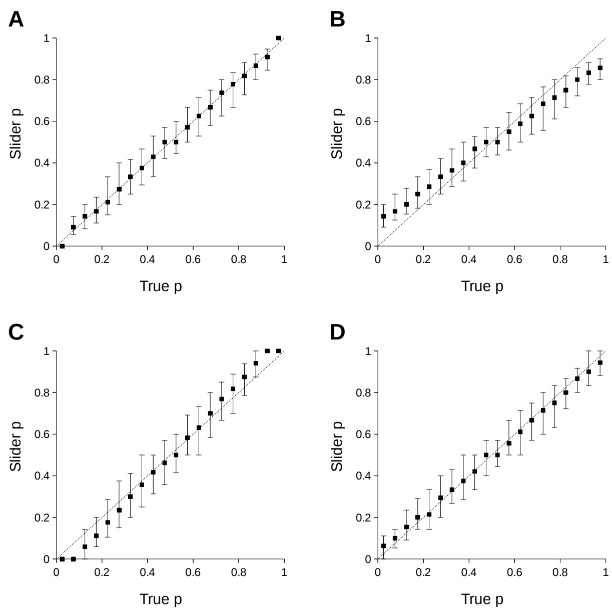

Figure 2. Median (squares) and interquartile intervals (vertical lines) of model's probability estimates plotted against corresponding true probabilities, for different values of the noise parameters: $d = 0.0, d_s = 0.0$(graph A), $d = 0.1, d_s = 0.0$ (graph B) $d = 0.0, d_s = 0.1$ (graph C) and $d = 0.1, d_s = 0.1$ (graph D). The dashed line represents identity.

forms of estimation, these contrasting effects of noise cancel out, producing agreement with the true probability. To test these predictions, for each simulation we calculated the median estimate produced by the model for a given hidden probability value p. The results are shown in the 4 graphs in Figure 2. Graph A gives the results obtained when there is no noise in either descriptive estimation or inferential judgment ($d = 0.0, d_s = 0.0$); the relationship between median estimates and the true probability is one of identity here. Graph B gives the results with noise in descriptive estimation but not inferential judgment ($d = 0.1, d_s = 0.0$), and shows a clear pattern of regression. Graph C gives the results with no noise in descriptive estimation but noise in inferential judgment ($d = 0.0, d_s = 0.1$), and shows a clear anti-regressive pattern. Finally, graph D shows the results obtained when there is the same rate of noise in both components ($d = 0.1, d_s = 0.1$). The relationship between median estimates and the true probability in graph D is one of identity: the effects of noise in the two components have cancelled each other out.

These results show that, if we assume a constant rate of error $d = 0.1$ in both descriptive probability estimation and inferential probability judgment, the probability theory plus noise model produces results that agree closely with those seen in Gallistel et al. (2014). Similar agreement holds for a range of other values of d. These same values of d, however, also produce regressive effects; in our model these regressive effects produce patterns of bias such as conservatism, subadditivity and the conjunction fallacy. In other words, this model may provide a single unified account for systematic bias away from the true probabilies (in some tasks) and for agreement with the true probabilities (in other tasks): an account that depends on a single factor - noise in reasoning.

Conclusions

Our aim in this paper is to present a general model of descriptive probability estimation, of inferential probability judgment, and of the interation between these two processes. This model assumes that people estimate (descriptive and inferential) probabilities using a mechanism that follows standard frequentist probability theory, but is subject to the biasing effects of random noise in the reasoning process. In other work we've shown that this model makes a number of novel predictions about patterns of bias and agreement with probability theory for various probabilistic expressions: predictions which are strongly supported by experimental results (see Costello and Watts, 2016a, 2014, 2016b). Here we show that this model can simultaneously explain the observed patterns of bias seen in people's descriptive probability estimation and inferential probability judgment (which arise in the model due to the regressive effects of random noise), and the observed agreement with the underlying true probability in tasks such as that of Gallistel et al.'s (where the regressive effect of noise in descriptive probability estimation is counteracted by the anti-regressive effect of noise in inferential probability judgment).

References

Brunswik, E. (1955). In defense of probabilistic functionalism: A reply. *Psychological Review*, 62:236–242.

Costello, F. and Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, 121(3):463–480.

Costello, F. and Watts, P. (2016a). Explaining high conjunction fallacy rates: the probability theory plus noise account. *Journal of Behavioral Decision Making*. In press, available at http://dx.doi.org/10.1002/bdm.1936.

Costello, F. and Watts, P. (2016b). People's conditional probability judgments follow probability theory (plus noise). *Cognitive Psychology*, 89:106–133.

Erev, I., Wallsten, T. S., and Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, 101(3):519–527.

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., and Latham, P. E. (2014). The perception of probability. *Psychological Review*, 121(1):96.

Peterson, C. and Beach, L. (1967). Man as an intuitive statistician. *Psychonomic Bulletin*, 68(1):29–46.

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315.

Tversky, A. and Koehler, D. J. (1994). Support theory: a nonextensional representation of subjective probability. *Psychological Review*, 101(4):547.