

9.07 Introduction to Statistics for Brain and Cognitive Sciences
Emery N. Brown

Lecture 2 Discrete Probability Models

I. Objectives

Introduce the concept of a random variable

Introduce the 4 principal discrete probability models

- a. Bernoulli
- b. Binomial
- c. Poisson
- d. Discrete counting probability model

Understand their basic assumptions, properties and situations for which they would be appropriate models

Understand the concepts of expectation and variance for discrete random variables

Definition 2.1. A **random variable** is a real-valued function from an outcome space into the real line more generally into R^n . The probability law defined on the outcome space induces (defines explicitly) a probability model for the random variable. A random variable can be either discrete or continuous. It is the basic quantity used in probability theory to characterize a probability process.

Example 2.0 Sum of the Digits on Two Fair Dice. To motivate the concept of a random variable we consider a single roll of two fair dice. In this case there are 36 outcomes because by the **Multiplication Principle (Definition 1.3)** we multiply the 6 possible outcomes on one die by the six possible outcomes on the second die. These outcomes are shown below in Table 1.

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Table 1. Outcomes from the roll of two fair dice.

Let X be the random variable that is defined as the sum of two fair dice. All the like values of X can be easily identified by looking along the diagonals of **Table 1** from right to left. By counting the number of distinct values on these diagonals we see that there are 11 possible outcomes. The probability of each of the 11 outcomes can be computed as

$$\Pr\{X = 2\} = \Pr\{(1,1)\} = \frac{1}{36}$$

$$\Pr\{X = 3\} = \Pr\{(1,2), (2,1)\} = \frac{2}{36}$$

$$\Pr\{X = 4\} = \Pr\{(1,3), (2,2), (3,1)\} = \frac{3}{36}$$

$$\Pr\{X = 5\} = \Pr\{(1,4), (2,3), (3,2), (4,1)\} = \frac{4}{36}$$

$$\Pr\{X = 6\} = \Pr\{(1,5), (2,4), (4,2), (3,3), (5,1)\} = \frac{5}{36}$$

$$\Pr\{X = 7\} = \Pr\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} = \frac{6}{36}$$

$$\Pr\{X = 8\} = \Pr\{(2,6), (3,5), (4,4), (5,3), (6,2)\} = \frac{5}{36}$$

$$\Pr\{X = 9\} = \Pr\{(3,6), (4,5), (5,4), (6,3)\} = \frac{4}{36}$$

$$\Pr\{X = 10\} = \Pr\{(4,6), (5,5), (6,4)\} = \frac{3}{36}$$

$$\Pr\{x = 11\} = \Pr\{(5,6), (6,5)\} = \frac{2}{36}$$

$$\Pr\{X = 12\} = \Pr\{(6,6)\} = \frac{1}{36}$$

We need a rule or summary process to describe the likelihood, frequency or probability with which a random variable assumes a set of values. In what follows, we will in general not talk about outcome spaces but instead in terms of random variables and the probability models that define the behavior of these random variables.

III. Discrete Probability Models

We divide probability models into two classes: discrete data probability models and continuous probability models. Discrete data are data which can assume a finite or a countably infinite set of values. In this lecture we discuss four discrete probability models: the Bernoulli, the binomial, the Poisson and a discrete counting probability model.

A. Bernoulli Probability Model

Example 2.1 (Learning Experiment, Jog et al. 1999)

To help establish neural correlates of procedural learning, Ann Graybiel and colleagues recorded from neurons in the striatum of rats over several days as they executed a procedure learning task. In this task, the rat used auditory cues to learn which one of two arms of a T-maze to enter in order to receive a reward. On each trial, the rat was placed in a T-maze. A tone was played. If it was a low tone the animal had to go left to receive a reward, whereas if it was a high tone it had to go right to obtain its reward. Suppose that on the previous day, the animal executed this task 40 times and, in so doing, made 22 correct choices and 18 incorrect choices. Before the start of the 40 trials today, what is the probability that the animal will give a correct response on a given trial?

In this problem, there are only two possible outcomes: a correct response or an incorrect response. The outcomes are mutually exclusive. That is, when one outcome occurs, the other cannot occur. Let p be the probability of a correct response, then $1-p$ is the probability of an incorrect response.

We can define this in terms of a random variable as follows. Let X be the random variable that is 1 if the response is correct and 0 if the response is incorrect. We can write

$$\begin{aligned} \Pr(X = 0) &= 1 - p \\ \Pr(X = 1) &= p \end{aligned} \tag{2.1}$$

or, in a more compact notation,

$$\Pr(X = x) = \Pr(X = 0 \text{ or } X = 1) = p^x(1-p)^{1-x}. \tag{2.2}$$

Equation 2.2 defines the Bernoulli probability model. It is the simplest probability model possible as there are only two outcomes. It is used extensively to model binary outcomes, e.g. yes-no, correct-incorrect, success-failure and spike-no spike type events. If we were to plot Eq. 2.2, we get

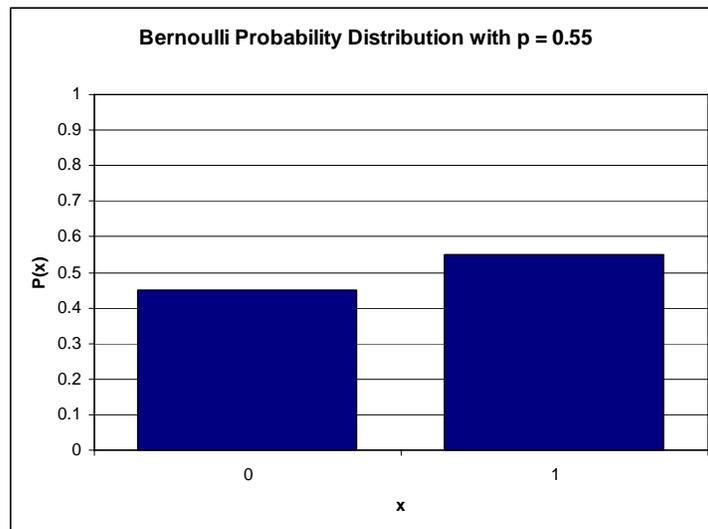


Figure 2A. Bernoulli Probability Mass Function with $p = 0.55$.

The **probability mass function** (pmf) defines the probability of each outcome. Notice that by **Definition 2.1** the sum of the probabilities of all the outcomes must be 1. Formally, we state this as

$$\sum_{x=0}^1 p(X = x) = (1 - p) + p = 1. \quad (2.4)$$

The **cumulative distribution function** (cdf) defines $F(x) = \Pr(X \leq x)$. We have $F(0) = \Pr(X \leq 0) = \Pr(X = 0) = 1 - p$ and $F(1) = \Pr(X \leq 1) = \Pr(X = 0) + \Pr(X = 1) = (1 - p) + p = 1$. Alternatively,

$$F(0) = \Pr(X \leq 0) = \sum_{x=-\infty}^0 p(x) = 1 - p \quad (2.5)$$

$$F(1) = \Pr(X \leq 1) = \sum_{n=0}^1 p(n) + p(1) = (1 - p) + p = 1. \quad (2.6)$$

The cdf at $F(X)$ defines the area under the curve up to and including X

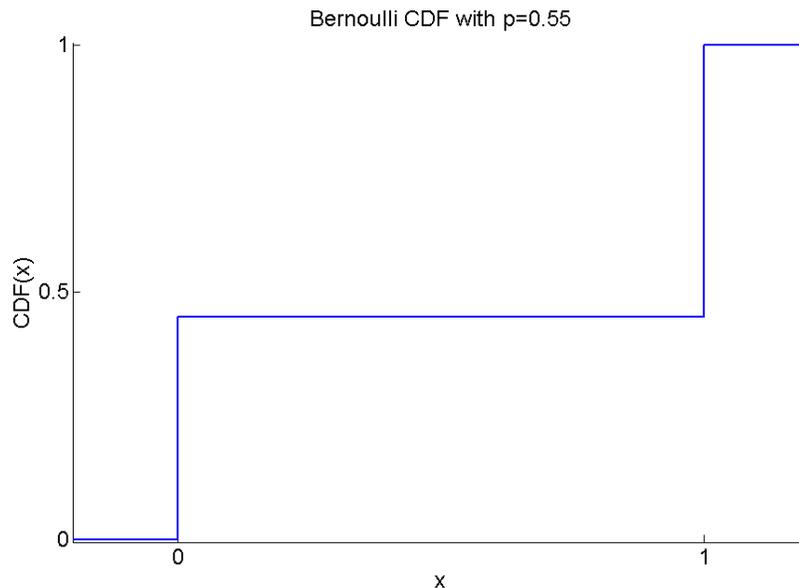


Figure 2B. Bernoulli Cumulative Distribution Function with p = 0.55.

Example 2.1 (continued). The Bernoulli probability model would be a good model for this problem. It defines the probability of a correct or incorrect response on each trial. Either the quantity $\hat{p} = \frac{22}{40} = 0.55$ or $p = 0.5$ would be a reasonable guess (or estimate) of p for a trial in today's experiment. Use of the parameter $\hat{p} = 0.55$ would suggest that we expect the

performance on the trials today to be like the performance on trials yesterday. Use of $p = 0.5$ as the guess of p for today would indicate a belief that performance will be indistinguishable from chance on today's trial.

Other theoretical properties of a probability model to characterize are the **mean** or **expected value** and the **variance**. The mean or average is defined as

$$\mu = E(X) = \sum_{x=-\infty}^{\infty} xp(x) \quad (2.7)$$

and the variance is

$$\begin{aligned} \sigma^2 &= \sum_{x=-\infty}^{\infty} (x - E(X))^2 p(x) \\ &= E[X - E(X)]^2 = E[X - \mu]^2 \\ &= E(X^2) - \mu^2. \end{aligned} \quad (2.8)$$

For our Bernoulli probability model we have

$$\begin{aligned} \mu = E(X) &= \sum_{x=0}^1 xp(x) \\ &= 0 \times (1-p) + 1 \times p \\ &= p \end{aligned} \quad (2.9)$$

$$\begin{aligned} \sigma^2 = E(X^2) - \mu^2 &= \sum_{x=0}^1 x^2 p(x) - p^2 \\ &= 0^2(1-p) + 1^2 p - p^2 \\ &= p - p^2 = p(1-p) \end{aligned} \quad (2.10)$$

The mean is the average value of the outcomes whereas the variance defines the spread of the pmf. For the Bernoulli probability model, note that if $p < \frac{1}{2}$ then the success or correct response with probability p is not **the most likely outcome** or the **mode**, which in this case would be the failure or the incorrect response whose probability is $1-p$. The variance for the Bernoulli random variable is $p(1-p)$. If p is small compared with 1, then $p(1-p) \approx p$. Hence, for small probability of a correct response, the mean and the variance of the Bernoulli are approximately equal. We will return to this observation when we discuss the Poisson probability model. The variance is a negative quadratic function of p (Figure 2C). This function has a maximum at $p = \frac{1}{2}$, where it has the value $\frac{1}{4}$. This makes the intuitive statement that probabilities close to $\frac{1}{2}$ will be most variable, i.e., have the highest mixture of success and failure, whereas probabilities close to either 0 or 1 will have to be less variable and have respectively predominantly either failures or successes respectively.

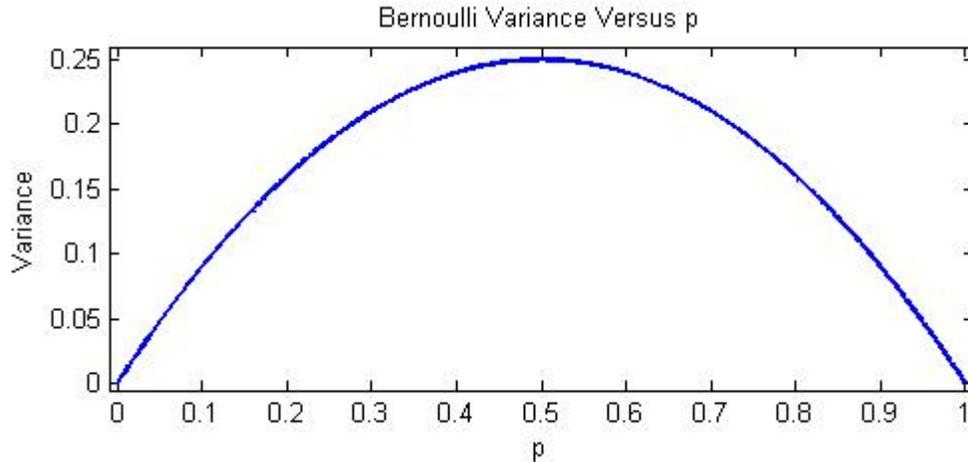


Figure 2C. Bernoulli Variance as a Function of p .

B. Binomial Probability Model

Example 2.1 (continued). If today we test the rat on the same task and give it 40 trials, and we assume that the probability of a correct choice is p , what are the possible outcomes and what is the probability of each outcome?

To answer this we must count the possible outcomes and assign each outcome the appropriate probability. Before beginning this computation we restate two results from **Lecture 1** that will be crucial. If a set of events E_1, \dots, E_n is **independent**, then

$$\begin{aligned} \Pr(E_1 \cap E_2 \dots \cap E_n) &= \Pr(E_1) \times \Pr(E_2) \times \dots \times \Pr(E_n) \\ &= \prod_{i=1}^n \Pr(E_i). \end{aligned} \quad (2.11)$$

That is, if a set of events is independent the probability of the joint or simultaneous occurrence of the events can be computed as the product of the probabilities of the individual events (See **Proposition 1.6** for a proof of this result). If a set of events E_1, \dots, E_n is mutually exclusive, then

$$\Pr(E_1 \cup E_2 \cup \dots \cup E_n) = \sum_{i=1}^n \Pr(E_i). \quad (2.12)$$

This results states that if a set of events is mutually exclusive the probability of the union is the sum of the probabilities. This result is a special case of third **Axiom of Probability** in **Lecture 1**.

On each trial, there can be either a correct or an incorrect response. Across the 40 trials, there can be any combination of correct and incorrect responses such that the sum of correct and incorrect responses equals 40. That is, there are k correct responses and $40-k$ incorrect responses for $k = 0, \dots, 40$. If we assume that the trials are independent, and that on each trial the probability of a correct response is p and the probability of an incorrect response is $1-p$, then for any sequence of k correct responses and $40-k$ incorrect responses the probability of this sequence or event is

$$\Pr(\text{any } k \text{ correct} | 40 \text{ trials}) = p^k (1-p)^{40-k} \quad (2.13)$$

If there are k correct responses and $40 - k$ incorrect responses then, there are $\binom{40}{k} = \frac{40!}{k!(40-k)!}$ combinations of such responses. The quantity $\binom{40}{k}$ is the number of ways of choosing k objects from 40 without regard to order and $k! = k \times (k-1) \times (k-2) \times \dots \times 2 \times 1$. For $k = 0, 1, 2, \dots, 40$, there are $\binom{40}{k}$ mutually exclusive sequences hence, by Eq. 2.12, we have the desired probability as

$$\Pr(k \text{ correct} | 40 \text{ trials}) = \binom{40}{k} p^k (1-p)^{40-k}. \quad (2.14)$$

Equation 2.14 is the binomial pmf with number of trials equal to 40 and probability of a correct response p . In general, we have for the binomial pmf

$$\Pr(k \text{ correct} | n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k} \quad (2.15)$$

$k = 0, 1, 2, \dots, n$. To see that Eq. 2.15 is a pmf note that

$$\sum_{k=0}^n \Pr(k | n) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1 \quad (2.16)$$

by the **Binomial Theorem (Eq. 1.11)** The cdf in this case is

$$F(k) = \sum_{j=0}^k \Pr(j | n) = \sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} \quad (2.17)$$

for $k = 0, 1, 2, \dots, n$. The pmf and the cdf for the learning experiment with 40 trials are shown below for $n = 40$ and $p = 0.55$. Hence, Eq. 2.15 defines the pmf of the **binomial probability model** with n trials and probability of a correct response or a success of p . Eq. 2.17 defines the cdf for the binomial probability model.

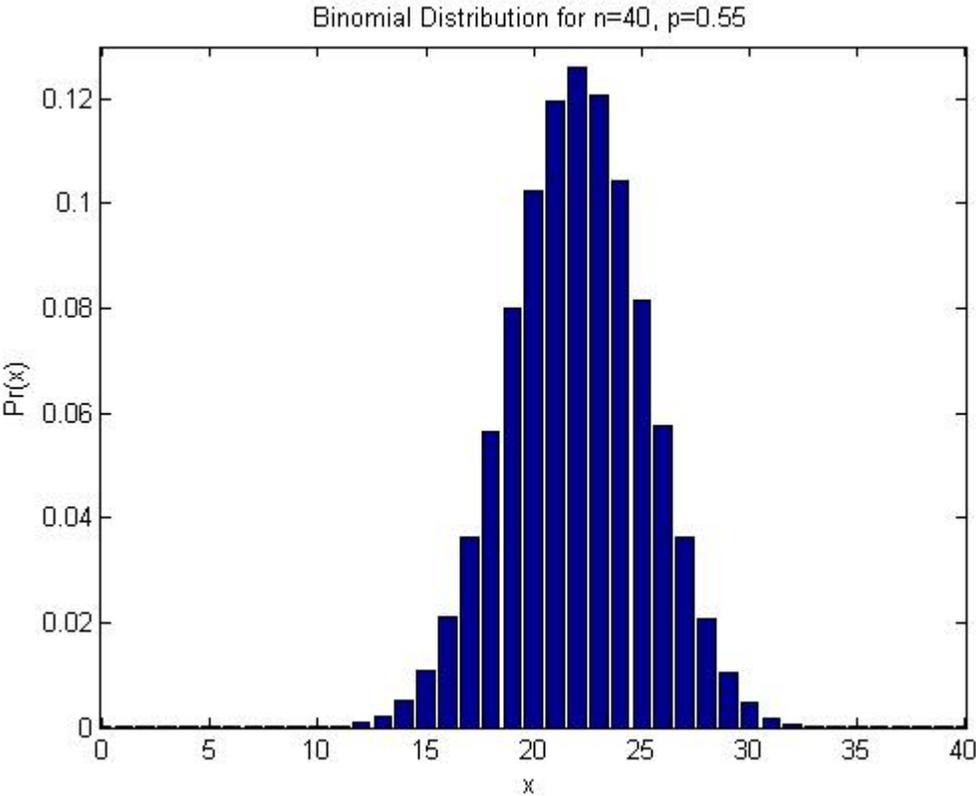


Figure 2D. Binomial Probability Mass Function for $n = 40$ and $p = 0.55$.

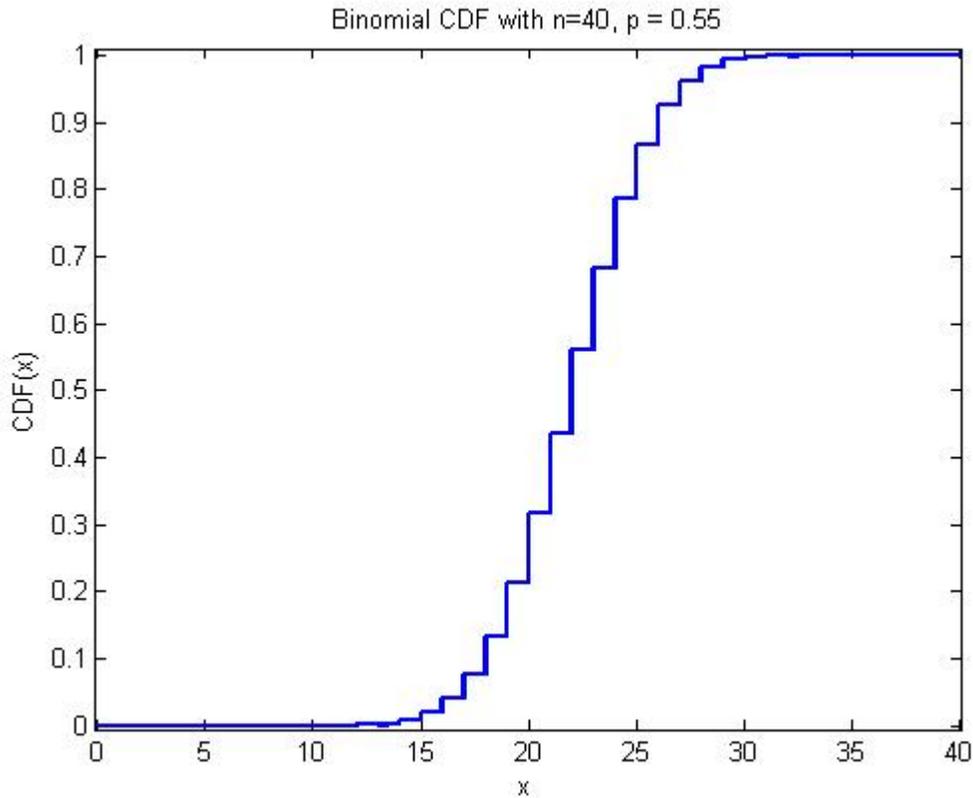


Figure 2E. Binomial Cumulative Distribution Function for $n = 40$ and $p = 0.55$.

Question 2.1. How does the outcome space of the binomial probability model with 40 trials relate to the outcome space of the Bernoulli probability model with 1 trial?

Example 2.1 (continued). Today, the rat executes 36 trials correctly out of 40. How likely is it to record 36 correct trials out of 40 if the probability of a correct response is the probability of a correct response determined from the animal's performance on the previous day $\hat{p} = 0.55$? If the new \hat{p} is $\frac{36}{40}$, is there a difference in performance between yesterday and today? To answer the first question we compute using Eq. 2.13

$$\begin{aligned}
 \Pr(k = 36 \mid \hat{p} = 0.55, n = 40) &= \binom{40}{36} (\hat{p})^{36} (1 - \hat{p})^4 \\
 &= \binom{40}{36} (0.55)^{36} (0.45)^4 \\
 &= 0.0000016857980774740635.
 \end{aligned}
 \tag{2.18}$$

Hence, it is highly unlikely to obtain 36 of 40 correct responses if the true probability is 0.55. Furthermore, it is even less unlikely if we assume that the animal is performing at chance because the probability of a correct response is 0.50. This would suggest that the animal is not performing as on the previous day or by chance. Furthermore, because $36/40 = 0.90$ is larger than both 0.50 and 0.55, this analysis suggests that the animal may have learned. The first

question foreshadows our study of hypothesis testing in **Lecture 12**. We will return to the second question when we learn how to compute confidence intervals.

The independence assumption was essential for these computations. It would not be appropriate if the animal had a tendency to move in one direction more than another, or if the animal's performance on a given trial began to depend on its performance on the previous trial or sequence of previous trials (Smith et al. 2004).

Other Theoretical Properties of the Binomial Probability Model

For the Bernoulli probability model it is easy to show that the expected value and variance are

$$\begin{aligned}\mu &= E(k) = np \\ \sigma^2 &= \text{Var}(k) = np(1-p)\end{aligned}\tag{2.19}$$

To compute the expected value, we apply the definition in Eq. 2.7 and obtain

$$\begin{aligned}E[k] &= \sum_{k=0}^n kp(k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ &= np[p + (1-p)]^{n-1} = np\end{aligned}$$

Providing the result for the variance directly from the definition is tedious!! We will develop alternative approaches in **Lecture 6**.

C. Poisson Probability Model

Example 2.2. The Quantal Release Hypothesis. Bernard Katz and colleagues (Tuckwell, 1988; Aidley, 1988) formulated the quantal hypothesis for the release of acetylcholine at the frog motor neuromuscular junction. It stated that in response to stimulation, acetylcholine is released from the motor nerve terminal in discrete "packets" or quanta. Normal endplate potentials (EPPs) are the result of several hundred quanta. A miniature EPP's are the result of spontaneous release of a single quantum. An important corollary of the quantal release hypothesis is that there is most likely a large population of quanta in the nerve terminal, each one of which has a small probability of being released by a nerve impulse. We now know that these quanta are packaged in vesicles. For a fixed small time interval (fraction of a millisecond) can we compute the probability that a given number of quanta or vesicles will be released?

To study this problem we can formulate a binomial probability model in which n is the number of quanta or release sites and p is the probability of release in a given small time interval. Let us assume that the release sites behave independently. This then leads to the binomial probability model and hence the probability of observing exactly k quanta released in the specified small time interval is given explicitly by Eq. 2.15. As n becomes large, it is more and more challenging to evaluate the factorials in Eq. 2.15. We can approximate this calculation by assuming that as n increases the probability of release decreases so that $n \times p \rightarrow \lambda$. Hence, for n sufficiently large, we have $np \approx \lambda$ or $p = \frac{\lambda}{n}$ and substituting into the binomial pmf, we obtain

$$\begin{aligned}
 \Pr(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \frac{n(n-1)(n-2), \dots, n-k+1}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{n(n-1)(n-2), \dots, n-k+1}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
 &\approx \frac{\lambda^k}{k!} e^{-\lambda}.
 \end{aligned} \tag{2.20}$$

These results follow because as $n \rightarrow \infty$

$$\begin{aligned}
 \frac{n(n-1)(n-2), \dots, n-k+1}{n^k} &\rightarrow 1 \\
 \left(1 - \frac{\lambda}{n}\right)^n &\rightarrow e^{-\lambda} \\
 \left(1 - \frac{\lambda}{n}\right)^{-k} &\rightarrow 1
 \end{aligned} \tag{2.21}$$

The last line of Eq. 2.20 states that the binomial pmf can be approximated by a Poisson pmf when n is large and p is small. The result is termed the **Poisson approximation to the binomial distribution**.

The **Poisson probability mass function** is defined formally as

$$\Pr(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \tag{2.22}$$

for $k = 0, 1, 2, 3, \dots$. This probability model is used to define the pmf of exactly k events occurring in a specified unit of time or space. The parameter λ is the rate parameters.

Other Properties of the Poisson Model

We have that if $X \sim P(\lambda)$ (read as “ X is distributed as a Poisson random variable with parameter λ ”) then

$$\begin{aligned}
 E(X) &= \lambda \\
 \text{Var}(X) &= \lambda
 \end{aligned} \tag{2.23}$$

Equation 2.23 is the well-known result that the mean and variance of a Poisson random variable are equal. These results are easy to establish directly from the definitions. Note that for the expected value we have

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} e^{-\lambda} \\
&= \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} e^{\lambda} = \lambda
\end{aligned} \tag{2.24}$$

where to go from line 2 to line 3 we made the change of variable $k = x - 1$.

Poisson Models in Perspective

The Poisson model is a central model in many theoretical analyses in computational neuroscience of neural spike trains (Rieke et al, 1997; Dayan and Abbott, 2001). While it is acknowledged that rarely do neural systems display Poisson behavior, this model is often used because it is very tractable analytically. Similarly, although the Poisson model came about as a natural extension of the binomial model for the analysis of the quantal release hypothesis, reports appeared within a short time of Katz's original papers to show that this model did not describe acetylcholine release accurately (Tuckwell, 1988).

The cases where good agreement between a Poisson model and an experimental system has been established are mostly empirical and not derived from first principles. Two of these are well known. Attention was significantly drawn to the probability model that is now called the Poisson distribution in 1898 in a paper by Ladislaus von Bortkiewicz. He noted how deaths of Prussian soldiers caused by horse kicks over a 20 year period in the 19th century could be described as a Poisson model and how their number could be estimated by the equation for the Poisson pmf (Rice 2007). Lord Rutherford presented an empirical analysis to show that the scintillations emitted by the radioactive decay of polonium in a 125 msec time interval obeyed a Poisson distribution (Rutherford and Geiger, 1910).

Several authors have shown that the Poisson model is not an accurate description of neural spiking data (Kass and Ventura 2001; Barbieri et al. 2001). In our recent work we have looked specifically into establishing alternative models for neural data analysis that do not require Assumption 3 (Brown et al. 2002; Truccolo et al. 2005). The fact that plausible alternatives exist has been known for years (Gerstein and Mandelbrot, 1964; Brillinger 1988) but these alternatives have not been widely used in neuroscience data analysis. There are now sufficiently many alternatives to the Poisson model that may be used in neural data analyses.

D. Discrete Probability Models

Any function defined on a finite set of values or any function defined on a countably infinite set of values such that the sum of the function of all the values is finite can be converted into a pmf for the set of values. This is because by assumption we have

$$\sum_{x=0}^{\infty} g(x) = c < \infty \tag{2.32}$$

and hence we can define

$$p(x) = c^{-1}g(x) \tag{2.33}$$

is a pmf defined on $x = 0, 1, 2, \dots$. The mean and variance are computed as in Eqs. 2.7 and 2.8 respectively. Hence, any finitely summable function can be converted into a pmf.

Example 2.5. Birthdays of Students in 9.07 in Spring 2007. Paymon collected the birthdays from 46 of the students in the 9.07 in the spring of 2007 and made a histogram of the data plotted by month. The histogram $g(x)$ is shown in Figure 2F panel 1 whereas the pmf $p(x)$ of the data is shown in panel 2. This $x = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$. An exploratory data analysis histogram or stem-and-leaf plot of the data is shown panel 3. Note that even from the small sample there is a suggestion that the birthdays are distributed uniformly across the 12 months.

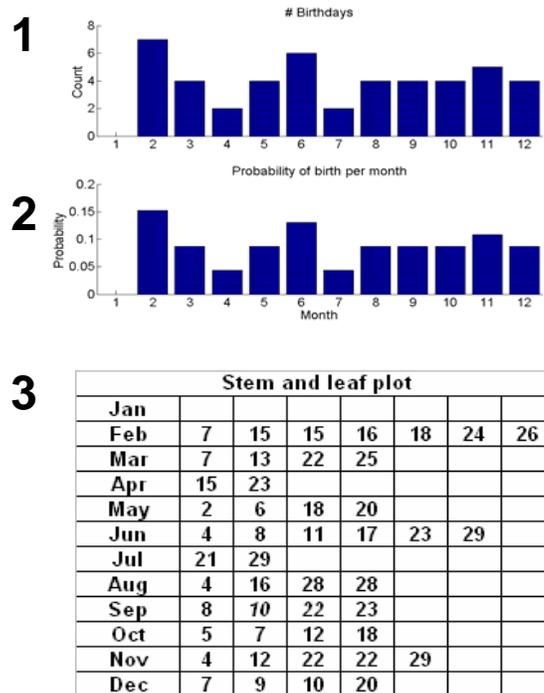


Figure 2F. Construction of a pmf on a set of discrete values using a sample of the birthdays of the students in 9.07. Panel 1 is the histogram of the birthdays. Panel 2 is the pmf and Panel 3 is the associated stem-and-leaf plot.

IV. Summary

We have introduced three discrete probability models, the Bernoulli, the binomial and the Poisson for our statistical analyses of neuroscience data. We discussed the models and their properties in relation to actual neuroscience experiments in which they were applied. These models will be useful in their own right and will serve as building blocks for more detailed data models.

Acknowledgments

I am grateful to Uri Eden for making Figures 2A-E, to Paymon Hosseini for making Figure 2F, to Jim Mutch for editorial comments and to Julie Scott for technical assistance.

Text References

Ross SM. *Introduction to Probability Models*. London, UK: Academic Press, 1993.

Dayan P. & Abbott, L.F. *Theoretical Neuroscience*. (MIT Press, Cambridge, 2001).

Rice, JA, *Mathematical Statistics and Data Analysis*, 3rd edition. Boston, MA, 2007.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. *Spikes Exploring the Neural Code* (MIT Press, Cambridge, 1997).

Literature References

Aidley DJ. *The Physiology of Excitable Cells*, 4th Edition. Cambridge: Cambridge University Press, 1998.

Barbieri, R., Quirk, M.C., Frank, L.M., Wilson, M.A. & Brown, E.N. Construction and analysis of non-Poisson stimulus response models of neural spike train activity. *J. Neurosci. Meth.* 105, 25-37 (2001).

Brillinger DR, Maximum likelihood analysis of spike trains of interacting nerve cells, *Biological Cybernetics*, Vol. 59, pp. 189-200, 1988.

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells, *Journal of Neuroscience*, 18:7411-25, 1998.

Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 2002, 14(2):325-46.

Gerstein GL, Mandelbrot M. Random walk models for the spike activity of a single neuron. *Biophysical Journal* 4: 41-68, 1964.

Jog MS, Kubota Y, Connolly CI, Hillegaard V, Graybiel AM Building neural representations of habits. *Science* 286:1745-1749, 1999.

Kass RE, Ventura V. A spike train probability model, *Neural Computation*, 13: 1713-1720, 2001.

Rutherford E Geiger H, The probability variations in the distribution of α particles. *Philosophical Magazine*, Sixth Ser., 20, 121-128, 1910.

Truccolo W, Eden UT, Fellow M, Donoghue JD, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble and covariate effects. *Journal of Neurophysiology*, 93:1074-1089, 2005

Tuckwell, HC. *Introduction to Theoretical Neurobiology: Nonlinear and Stochastic Theories, Volume 2*. Cambridge: Cambridge University Press, 1988.