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 outlcomes on one die by the six possible outlcomes 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 dices.

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 models and continuous-valued data models. As stated in the **Introductory Lecture,** 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 is 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

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

or, in a more compact notation,

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

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

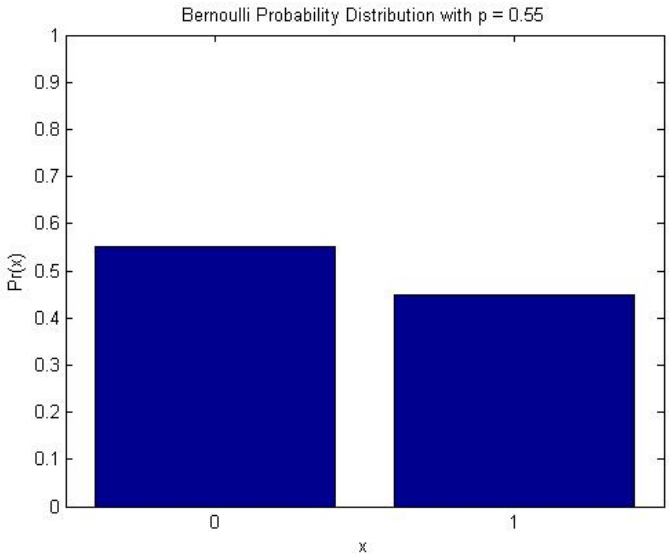

Figure 2A. Bernoulli Probability Mass Function with p = 0.55.

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

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

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

$$F(0) = \Pr(X \le 0) = \sum_{x=0}^{1} p(x) = p \tag{2.5}$$

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

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

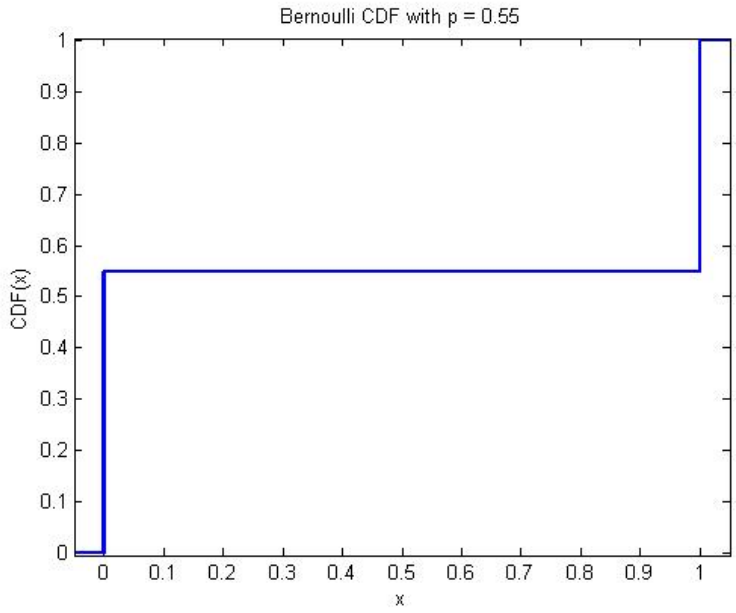

Figure 2B. Bernoulli Cumulative Distribution Function with p = 0.55.

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 estimate (guess) 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

$$\mu = E(X) = \sum_{x=0}^{\infty} xp(x) \tag{2.7}$$

and the variance is

$$\begin{aligned} \sigma^2 &= \sum_{x=0}^{\infty} (x - E(X))^2 p(x) \\ &= E[X - E(X)]^2 = E[X - \mu]^2 \\ &= E(X^2) - \mu^2. \end{aligned} \tag{2.8}$$

For our Bernoulli probability model we have

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

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

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 p is not **the most likely outcome** or the **mode**, which in this case would be $1 - p$. 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 success, 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.

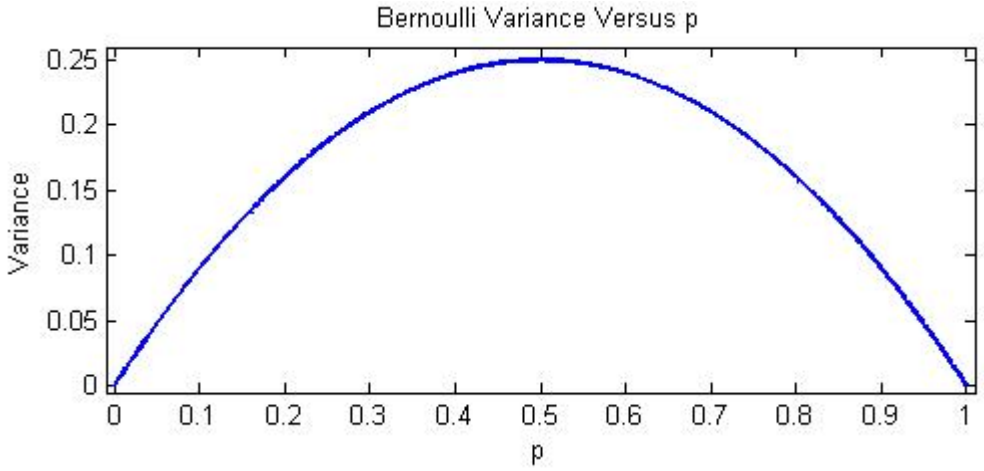

Figure 2C. Bernoulli Variance as a Function of p.

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 state two results from probability theory that will be useful. If a set of events $E_1, \ldots, E_n$ is **independent** then

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

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 $E_1, \ldots, E_n$ is mutually exclusive then

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

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 $k = 0, \ldots, 40.$ 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 if there are k correct responses and $40-k$ incorrect responses then the probability of this event is

$$\Pr(k \text{ successes} | 40 \text{ trials}) = p^k (1-p)^{40-k} \tag{2.13}$$

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 $k!=k\times(k-1)\times(k-2)\times...\times 2\times 1.$ For $k=0,1,2,...,N,$ there are $\binom{40}{k}$ mutually exclusive sequences hence, by Eq. 2.12, we have the desired probability as

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

Equation 2.12 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 is

$$\Pr(k\ \text{success}|N)=\binom{N}{k}p^k(1-p)^{N-k} \tag{2.15}$$

$k=0,1,2,...,N.$ To see that Eq. 2.13 is a pmf note that

$$\sum_{k=0}^{N}\Pr(k\mid N)=\sum_{k=0}^{N}\binom{N}{k}p^k(1-p)^{N-k}=1 \tag{2.16}$$

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

$$F(k)=\sum_{j=0}^{k}\Pr(j\mid N)=\sum_{j=0}^{k}\binom{N}{k}p^j(1-p)^{N-j} \tag{2.17}$$

for $k=0,1,2,...,N.$ The pmf and the cdf for the learning experiment with 40 trials are shown below for $N=40$ and $p=0.55.$

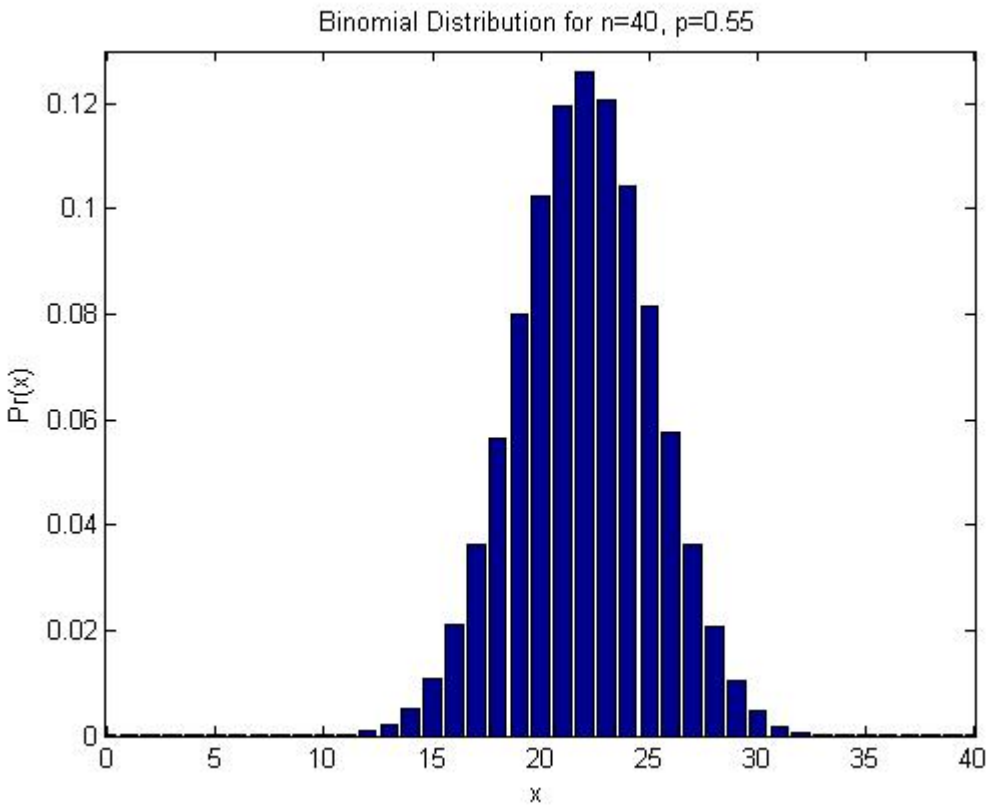

Figure 2D. Binomial Probability Mass Function for N= 40 and p = 0.55.

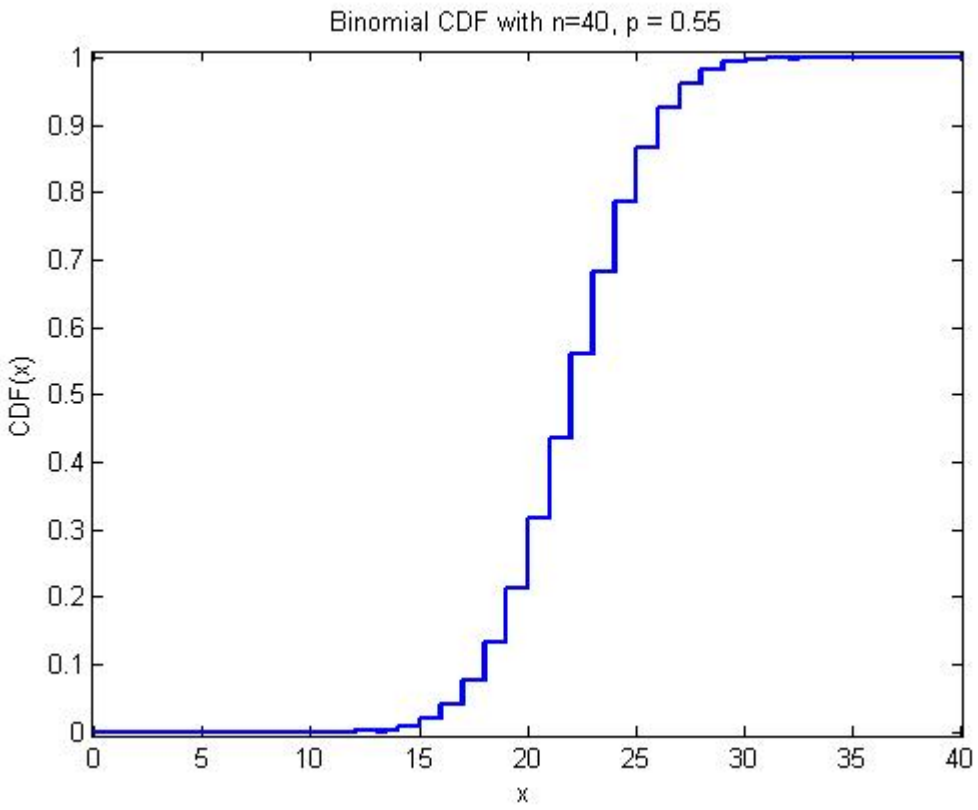

Figure 2E. Binomial Cumulative Distribution Function for N = 40 and p = 0.55.

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. 2.12

$$\begin{aligned} \Pr(k = 36 \mid \hat{p} = 0.55, 40) &= \binom{40}{36} (\hat{p})^{36} (1 - \hat{p})^{4} \\ &= \binom{40}{36} (0.55)^{36} (0.45).^{4} \\ &= 0.0.00000168579807747406 35. \end{aligned} \tag{2.18}$$

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 and because 0.90 is larger than both 0.50 and 0.55, this analysis suggests that the animal may have learned. 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.

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

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

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

$$E[X] = \sum_{k=0}^{N} kp(k) = \sum_{k=0}^{N} k\binom{N}{k} p^k (1-p)^{N-k} = Np\sum_{k=0}^{N-1}\binom{N-1}{k} p^k (1-p)^{N-1-k}$$

$$= Np[p+(1-p)]^{N-1} = Np$$

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. Miniature EPP's are the result of spontaneous release of single quanta. 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. 2.14. As a practical matter, as N becomes large, it is more and more challenging to evaluate the factorials in Eq. 2.14. We can approximate this calculation by assuming that as N increases the probability of release decreases so that $N \times p \to \lambda$. Hence, for N sufficiently large, we have $Np \approx \lambda$ or $p = \frac{\lambda}{N}$. and substituting into the binomial pmf, we obtain

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

These results follow because as $N \to \infty$

$$\begin{aligned}&\frac{N(N-1)(N-2),...,N-k+1}{N^k} \to 1 \\ &(1-\frac{\lambda}{N})^N \to e^{-\lambda} \\ &(1-\frac{\lambda}{N})^{-k} \to 1\end{aligned} \tag{2.21}$$

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 $k = 0,1,2,3,...$. 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 X is distributed as a Poisson random variable with parameter) then

$$\begin{aligned}&E(x) = \lambda \\ &Var(x) = \lambda \\ &\sigma_x = (Var(x))^{\frac{1}{2}} = \lambda^{\frac{1}{2}}\end{aligned} \tag{2.23}$$

These results are easy to establish. Note that for the expected value we have

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

The Poisson Assumptions and Poisson Model 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 analytically very tractable. 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 three Poisson assumptions define a Poisson process in that if a probability model satisfies these three assumptions, then it is a Poisson process. Assumptions 1 and 2 are reasonable for the analysis of neuronal data. The most challenging assumption is Assumption 3. It is difficult to believe that arbitrarily close non-overlapping time intervals are independent in a physical or biological system. 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. 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 (Truccolo et al. 2005). The fact that plausible alternatives exist has been known for year (Gerstein and Mandelbrot, 1964; Brillinger 1988) but these alternatives have not been widely applied. There are now sufficiently many alternatives to the Poisson model that maybe 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 $x = 0,1,2,\ldots$. 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 9.914 in Spring 2006. Uri collected the birthdays from 16 of the students in the class and made a histogram of the data plotted by month. The histogram $g(x)$ is shown in Figure 2H 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 EDA 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.

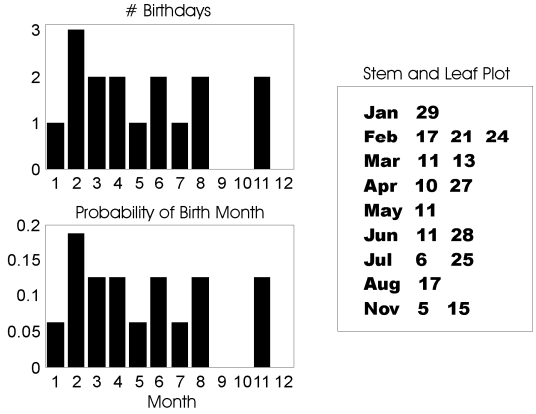

Figure 2H. Construction of a pmf on a set of discrete values using a sample of the birthdays of the students in 9.914. Panel 1 is the histogram of the birthdays. Panel 2 is the pmf and Panel 3 is the associated stem-and-leaf plot.

IV. Summary

We have developed a set of discrete probability models 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. The results in today's lecture fall into Step 3 (Models) of the Box-Tukey paradigm presented in Lecture 1.

V. Appendix

Poisson Process

A sequence of events $S_1, S_2, \ldots,$ that occur in an interval $(0,T]$ is called a Poisson process if in any infinitesimal interval $[t, t+\Delta)$ for t in $(0,T]$ the following conditions are satisfied

i) the probability of an event in $[t, t+\Delta)$ is $\lambda\Delta$.

ii) the probability of more than one event is much, much smaller than $\lambda\Delta$.

iii) if $[s, s+\Delta_s)$ and $[t, t+\Delta_t)$ are non-overlapping intervals in $(0,T]$, the probability of an event in the first interval is independent of the probability of an even in the second interval.

For a Poisson process the probability of x events in an interval of width Δ is

$$\Pr(X = x) = \frac{e^{-\lambda\Delta}(\lambda\Delta)^x}{x!} \tag{2.25}$$

which is Eq. 2.22 with parameter $\lambda = \lambda\Delta$. If we decompose time into a large number N of sufficiently small intervals Δ, we have effectively a sequence of N independent Bernoulli trials. This is because with Δ sufficiently small, there is a probability of $\lambda\Delta$ of an event and probability $1-\lambda\Delta$ of no event. We have from Eq. 2.25 that

$$\begin{aligned} \Pr(X = x) &= \frac{e^{-\lambda\Delta}(\lambda\Delta)^x}{x!} \\ &\approx \frac{(1-\lambda\Delta)^{1-x}(\lambda\Delta)^x}{x!} \\ &= (1-\lambda\Delta)^{1-x}(\lambda\Delta)^x \end{aligned} \tag{2.26}$$

We see that on a small time-scale the Poisson process is a Bernoulli random variable. This feature of the Poisson process is a feature of a more general class of discrete probability models called point processes. As we mentioned in Lecture 1, point processes are binary processes that occur in continuous time. A Poisson process is a point process. We will study point processes in detail in later lectures.

Inhomogeneous Poisson Process

If we let $\lambda = \lambda(t)$ be time-dependent, then we obtain an inhomogeneous Poisson process. In this case, the rate varies with time. If this process is defined on an interval $(0,T]$, then for any interval (t_1,t_2) in $(0,T]$ the rate function is

$$\Lambda(t_1,t_2) = \int_{t_1}^{t_2} \lambda(u)du \tag{2.27}$$

The associated inhomogeneous probability mass function is

$$\Pr(X(t_1,t_2) = k) = \frac{e^{-\Lambda(t_1,t_2)}\Lambda(t_1,t_2)^k}{k!} \tag{2.28}$$

for $k = 0,1,2,3,\ldots$. This probability model is used to define the pmf of exactly k events occurring in interval (t_1,t_2). The rate is now defined by the function $\Lambda(t_1,t_2)$. We have that if $X(t_1,t_2) \sim P(\Lambda(t_1,t_2))$ then

$$\begin{aligned} E(X(t_1,t_2)) &= (\Lambda(t_1,t_2)) \\ V(X(t_1,t_2)) &= (\Lambda(t_1,t_2)) \\ \sigma_{x(t_1,t_2)} &= [V(X(t_1,t_2)]^{\frac{1}{2}} = \Lambda(t_1,t_2)^{\frac{1}{2}} \end{aligned} \tag{2.29}$$

Technically speaking, the new random variable $X(t_1,t_2)$ is a stochastic process indexed by time. We will define this entity precisely when we study point processes and time-series in later

lectures. The Poisson process can also be distributed over space instead of time. In this case, it defines the number of events per unit space.

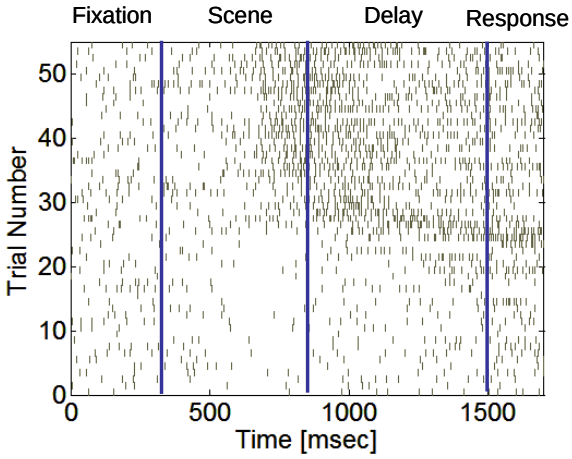

Figure 2F. Raster plot of single neuron from the monkey hippocampus during a 55 trial location-scene association task. Data courtesy of Wendy Suzuki NYU.

The Peristimulus Time Histogram (PSTH) is an example of an empirically constructed inhomogeneous Poisson model.

Example 2.3. Peristimulus Time Histogram (PSTH). Spike trains are often collected in response to a stimulus applied multiple times within a specified time interval. The stimulus may be explicit or implicit as discussed in Lecture One. Each application of a stimulus is called a trial and a standard way to present the data in either in a raster plot (Figure 2F) or as a peristimulus time histogram (PSTH) (Figure 2G). The raster is simply the plot of the spiking activity of a given neuron across its trials. The PSTH is computed from the raster plot binning the time axis and summing the number of spikes in a given bin across all trials. The PSTH is a much used technique in neural data analysis because it is easy to compute and it is often perceived as an analysis method that is assumption free. Clearly, this is not the case as the shape of the PSTH depends critically on the choice of bin size. Nothing about this choice is invariant.

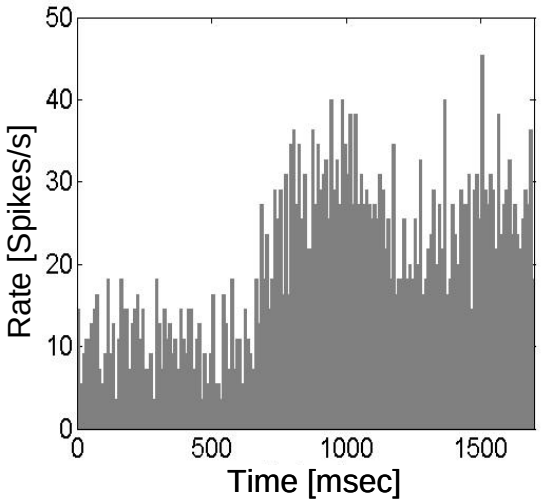

Figure 2G. Peristimulus Time Histogram of raster plot in Figure 2F. Data courtesy of Wendy Suzuki NYU.

Because the number of spikes per bin can often vary, the PSTH is often modeled as an inhomogeneous Poisson process. The argument to justify this assumption follows by applying the derivation of the Poisson approximation to the binomial separately to each bin across all trials. In this case the rate parameter varies as a function of the bin.

There has been theoretical work on general Poisson limit theorems. We mention one result that is relevant for neuroscience data analysis is in Brillinger (1976). Brillinger uses one to establish the large sample properties of his histogram-based estimate of the cross-intensity function for relating two neural spike trains. We will discuss the cross-intensity function later in our lectures on point processes. Brillinger shows that when the sample size is large adjacent bins in a histogram-based estimate of the cross-intensity function are independent Poisson random variables. A more careful analysis of the PSTH using this type of Poisson limit law may give a more rigorous theoretical justification for use of the Poisson model in PSTH analyses.

Example 2.4 Stimulus-Response Models and the Inhomogeneous Poisson Process. Other applications of the inhomogeneous Poisson model in neural spike train modeling include analysis of the spatial receptive fields of the rat hippocampus (Brown et al. 1998; Zhang et al. 1998; Barbieri et al. 2004) and modeling of MI spiking activity in association to hand position and or velocity (Brockwell et al. 2004; Wu et al. 2006; Truccolo et al. 2005). The appeal of the inhomogeneous Poisson model offers a very practical way to formulate a stimulus-response model because we can express neural spiking activity as a function of a time-varying covariate. Here are two examples.

Hippocampal Place Cell Model (Brown et al. 1998)
As a rat executes a behavioral task spiking activity of hippocampal neurons are known to have their spiking activity modulated by the animal's position in the environment and the phase of the theta rhythm. An inhomogeneous Poisson model for describing neural spiking activity as a function of the animal's position and the phase of the theta rhythm can be defined as

$$\lambda(t) = \exp(\alpha - \frac{1}{2}(x(t) - \mu)'W^{-1}(x(t) - \mu) + \beta\cos(\phi(t))), \tag{2.30}$$

where $\exp(\alpha)$ is the background firing rate, $x(t)$ is the position of the animal at time t, μ is the center of the place field, W is a scale matrix that governs the orientation of the field, $\phi(t)$ is the phase of the theta rhythm and β is the modulation parameter. Position and phase of the theta rhythm are the time-varying covariates.

Primary Motor Cortex Model (Moran and Schwarz 1999; Brockwell et al., 2004; Truccolo et al. 2005). Moran and Schwarz (1999) proposed a model that describes the spiking activity of MI neurons in terms of the velocity of a hand movement. Their description suggests a representation of the rate function for a Poisson model as

$$\lambda(t) = \exp(\alpha + \beta \mid v(t+\tau) \mid \cos(\phi(t+\tau)) + \gamma \mid v(t+\tau) \mid \sin(\phi(t+\tau))), \tag{2.31}$$

where $\exp(\alpha)$ is the background firing rate, $|v(t)|$ is the speed of the movement at time t, $\phi(t)$ is the direction of the movement at time t, τ is a lead time (negative delay) parameter, and β and γ are the modulation parameters. Speed and direction are the time-varying covariates.

Acknowledgments
I am grateful to Uri Eden for making Figures 2A-E and 2H, to Gabriela Czanner for making Figures 2F and 2G 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).

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).

Barbieri R, Frank LM, Nguyen DP, Quirk MC, Solo V, Wilson MA, Brown EN. Dynamic analyses of information encoding by neural ensembles. *Neural Computation*, 16 (2): 277-307, 2004.

Brillinger DR, Estimation of the second-order intensities of a bivariate stationary point process, *Journal of the Royal Statistical Society B* Vol. 38, pp. 60-66, 1976.

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

Brockwell, A.E., Rojas, A.L., and Kass, R.E. Recursive Bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology*, 91: 1899-1907, 2004.

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.

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, Hillegaart 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.

Moran DW, Schwartz AB. Motor cortical representations of speed and direction during reaching. *J. Neurophysiology.* 82:2676-2692, 1999.

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.