## Last Time

- Independence
- Bayes Theorem
- Classification
  - Naïve Bayes as a Classifier

1

## Today

- Discrete Random Variables
  - Probability Distributions
  - Probability Mass Functions
  - Cumulative Distribution Functions
  - Mean and Variance of a Discrete Random Variable
- Important Discrete Random Variables
  - Discrete Uniform Distribution
  - Binomial Distribution
  - Geometric Distribution
  - Poisson Distribution

2

# Probability Distributions

A random variable, X, is a function that assigns a real number, $\{x_1, x_2, \ldots, x_n\}$ to each outcome in the sample space of a random experiment.

The probability distribution of a random variable $X$ gives the probability for each value of $X$.

3

# Probability Mass Function

For a discrete random variable $X$ with possible values $x_1, x_2, \ldots, x_n$, a probability mass function is a function such that:

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$

4

## Credit Card Balance Payment

Monthly credit card debt payments are monitored for three consecutive periods. The probability that a payment is made on-time is 0.8, and the monthly payments are independent.

Let's answer the following:
- Sample Space: Discrete or continuous?
- List all possible outcomes
- Calculate the probabilities associated with outcomes
- What is the probability of getting at least two on-time payments?

Let the random variable *X* denotes the number of on-time payments. The last column of the table shows the values of *X* assigned to each outcome of the experiment.

| # Payment | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | Prob. | *X* |
| OK | OK | OK | 0.512 | 3 |
| OK | OK | nOK | 0.128 | 2 |
| OK | nOK | OK | 0.128 | 2 |
| OK | nOK | nOK | 0.032 | 1 |
| nOK | OK | OK | 0.128 | 2 |
| nOK | OK | nOK | 0.032 | 1 |
| nOK | nOK | OK | 0.032 | 1 |
| nOK | nOK | nOK | 0.008 | 0 |
| | | | 1.000 | |

| Probability Distribution of *X* | | | | |
|---|---|---|---|---|
| *P*(*X*=0) | *P*(*X*=1) | *P*(*X*=2) | *P*(*X*=3) | |
| 0.008 | 0.096 | 0.384 | 0.512 | 1 |



## Customer Life Time

- Let the random variable *X* denote the number of periods that a customer does business with the company. Assume that the probability that a customer defects with 0.01 probability in a given period; and that the behaviour over periods is independent. Determine the probability distribution of *X*.

Let «c» denotes a defecting event & let «a» denotes that the customer is in an active relationship with the company.
The sample space is: S = {c, ac, aac, aaac, …}
The range of the values of X is: x = 1, 2, 3, 4, …

| Probability Distribution | | |
|---|---|---|
| $P(X = 1) =$ | 0.01 | 0.01 |
| $P(X = 2) =$ | $(0.99)*0.01$ | 0.0099 |
| $P(X = 3) =$ | $(0.99)^2*0.01$ | 0.0098 |
| $P(X = 4) =$ | $(0.99)^3*0.01$ | 0.0097 |
| | | |

Question: What can you conclude about the following events?
= {acaa, caac, …}

## Cumulative Distribution Function and Properties

The cumulative distribution function, is the probability that a random variable *X* with a given probability distribution will be found at a value less than or equal to *x*.

Symbolically,

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

For a discrete random variable *X*, *F*(*x*) satisfies the following properties:

$$(1)\ F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

$$(2)\ 0 \le F(x) \le 1$$

$$(3)\ \text{If } x \le y, \text{ then } F(x) \le F(y)$$

7

## Cumulative Distribution Functions

Consider the probability distribution for the credit card payments

| Probability Distribution of *X* | | | | |
|---|---|---|---|---|
| *P*(*X*=0) | *P*(*X*=1) | *P*(*X*=2) | *P*(*X*=3) | |
| **0.008** | **0.096** | **0.384** | **0.512** | **1** |

Find the probability of two or fewer payments are on-time.

The event (X ≤ 2) is the total of the events: (X = 0), (X = 1) and (X = 2), From the table:

$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.488$

or, alternatively,

$P(X \le 2) = 1 - P(X = 3) = 1 - 0.512 = 0.488$

8

# Example: Sampling without Replacement

A day's production of 850 parts contains 50 defective parts. Two parts are selected at random without replacement. Let the random variable *X* equal the number of defective parts in the sample. Find the cumulative distribution function of *X*.

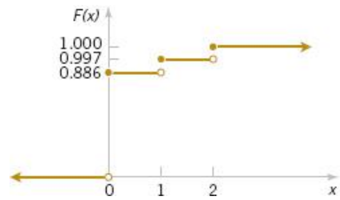
The probability mass function is calculated as follows:

$$P(X=0)=\tfrac{800}{850}\cdot\tfrac{799}{849}=0.886$$
$$P(X=1)=2\cdot\tfrac{800}{850}\cdot\tfrac{50}{849}=0.111$$
$$P(X=2)=\tfrac{50}{850}\cdot\tfrac{49}{849}=0.003$$

Therefore,

$$F(0)=P(X\le 0)=0.886$$
$$F(1)=P(X\le 1)=0.997$$
$$F(2)=P(X\le 2)=1.000$$

$$F(x)=\begin{cases}0 & x<0\\ 0.886 & 0\le x<1\\ 0.997 & 1\le x<2\\ 1 & 2\le x\end{cases}$$




Cumulative Distribution Function



# Expected Value of a Discrete Random Variable

- If *X* is a discrete random variable with probability mass function *f* (*x*), then the expected value, or mean, of X is defined as:

$$E[X]=\sum_x xf(x)$$

What is the expected value of the random variable X having the following probability distribution?

| Probability Distribution of *X* | | | | |
|---|---|---|---|---|
| *P(X=0)* | *P(X=1)* | *P(X=2)* | *P(X=3)* | |
| **0.008** | **0.096** | **0.384** | **0.512** | **1** |

$$E[X]=0(0.008)+1(0.096)+2(0.384)+3(0.512)=2.4$$

## Expected Value of a Function of a Discrete Random Variable

If $X$ is a discrete random variable with probability mass function $f(x)$,

$$E\left[h(X)\right] = \sum_x h(x) f(x)$$

The probability distribution of X is as follows. What is the expected value $X^2$?

| Probability Distribution of *X* | | | | | |
|---|---|---|---|---|---|
| ***X*** | **0** | **1** | **2** | ***3*** | ***4*** |
| **f(*X*)** | **0.6561** | **0.2916** | **0.0486** | **0.0036** | **0.0001** |

Here $h(X) = X^2$

$E(X^2) = X^2 \cdot f(X) = 0^2\,(0.6561) + 1^2\,(0.2916) + 2^2\,(0.0486) + 3^2\,(0.036) + 4^2\,(0.0001) = 0.5200$

Most of the time expected value of random variable is represented with te Greek letter $\mu$.

11

# Variance

$V(X) = \sum_x (x-\mu)^2 f(x)$ is the definitional formula

$$= \sum_x \left(x^2 - 2\mu x + \mu^2\right) f(x)$$

$$= \sum_x x^2 f(x) - 2\mu \sum xf(x) + \mu^2 \sum_x f(x)$$

$$= \sum_x x^2 f(x) - 2\mu^2 + \mu^2$$

$= \sum_x x^2 f(x) - \mu^2$ is the computational formula

Note: If $h(x) = (X-\mu)^2$, then its expectation is the variance of $X$.

$$V[X] = E[X^2] - E[X]^2$$

The standard deviation, σ ,is defined as the square root of the variance.

12

## Credit Card Example Revisited

| Probability Distribution of *X* | | | | |
|---|---|---|---|---|
| *P*(*X*=0) | *P*(*X*=1) | *P*(*X*=2) | *P*(*X*=3) | |
| **0.008** | **0.096** | **0.384** | **0.512** | **1** |

Calculate the mean & variance.

| X | f(x) | Xf(x) | (X-E[X])² | (X-E[X])²f(x) | X² | X²f(x) |
|---|---|---|---|---|---|---|
| 0 | 0,008 | 0,000 | 5,76 | 0,04608 | 0 | 0 |
| 1 | 0,096 | 0,096 | 1,96 | 0,18816 | 1 | 0,096 |
| 2 | 0,384 | 0,768 | 0,16 | 0,06144 | 4 | 1,536 |
| 3 | 0,512 | 1,536 | 0,36 | 0,18432 | 9 | 4,608 |
| | | 2,400 | | 0,480 | | 6,240 |
| | | | | | | |
| | E[X] | 2,400 | | | | |
| | V[X] | 0,480 | | | | |
| | | | | | | |



## Discrete Uniform Distribution

If the random variable $X$ assumes the values $x_1$, $x_2$, …, $x_n$, with equal probabilities, then the discrete uniform distribution is given by

$$f(x_i) = 1/n$$

## Discrete Uniform Distribution

- Let $X$ be a discrete random variable ranging from $a,a+1,a+2,\ldots,b,$ for $a \leq b$.
- There are $(b - a+1)$ values in the inclusive interval. Therefore:

$$f(x) = 1/(b-a+1)$$

- Its expected value and variance are as follows:

$$\mu = E(x) = (b+a)/2$$

$$\sigma^2 = V(x) = [(b-a+1)^2-1]/12$$

Note that the mean is the midpoint of $a$ & $b$.



## Dice

Let the random variable $X$ denote the outcome of throwing a dice. Can we assume that $X$ is a discrete uniform random variable? If so what is the range of $X$? Find $E(X)$ & $\sigma$.

$S(X) = \{1,2,3,4,5,6\}$

$E[X] = (6+1)/2 = 3.5$

$V[X] = 35/12$

$\sigma = (35/12)^{0.5}$

## Bernoulli Trial

A Bernoulli trial (or binomial trial) is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted.

Let $p$ be the probability of success in a Bernoulli trial, and $q$ be the probability of failure, such that $p + q = 1$

Note: A Binomial RV $X$ denotes the number of trials that result in a success for a sequence of bernoulli trials.



## Binomial Experiment: Biased Coin

- Let's say we are going to toss a biased coin three times. P(heads)=0.8 and P(tails)=0.2
- Can we assume that this is a binomial experiment?

  Yes, with success probability 0.8 or 0.2 depending on how we define success.

- Assume we toss the coin three times. What is the probability of getting E= {head, head, tail}?

  P(E ={heads, heads, tails}) = (0.8)(0.8)(0.2) = 0.128

- Assume we toss the coin three times. What is the probability of getting two heads and one tail?

  P({heads, heads, tails} + {heads, tails, heads} + {tails, heads, heads}) = 0.384

# Binomial Experiment: Biased Coin

| Heads # | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | Prob. | *X* |
| Heads | Heads | Heads | 0.512 | 3 |
| Heads | Heads | Tails | 0.128 | 2 |
| Heads | Tails | Heads | 0.128 | 2 |
| Heads | Tails | Fails | 0.032 | 1 |
| Tails | Heads | Heads | 0.128 | 2 |
| Tails | Heads | Fails | 0.032 | 1 |
| Tails | Tails | Heads | 0.032 | 1 |
| Tails | Tails | Tails | 0.008 | 0 |
| | | | 1.000 | |

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, \ldots n$$

| Probability Distribution of *X* | | | | |
|---|---|---|---|---|
| *P*(*X*=0) | *P*(*X*=1) | *P*(*X*=2) | *P*(*X*=3) | |
| 0.008 | 0.096 | 0.384 | 0.512 | 1 |

19

# Binomial Distribution

The random variable *X* that equals the number of trials that result in a success is a binomial random variable with parameters $0 < p < 1$ and $n = 1, 2, \ldots$.

The probability mass function is:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, \ldots n$$

For constants *a* and *b*, the binomial expansion is

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

20

# Binomial Coefficient

Exercises in binomial coefficient calculation:

$$\binom{10}{3} = \frac{10!}{3!7!} = \frac{10\cdot 9\cdot 8\cdot 7!}{3\cdot 2\cdot 1\cdot 7!} = 120$$

$$\binom{15}{10} = \frac{15!}{10!5!} = \frac{15\cdot 14\cdot 13\cdot 12\cdot 11.10!}{10!.5\cdot 4\cdot 3\cdot 2\cdot 1} = 3,003$$

$$\binom{100}{4} = \frac{100!}{4!96!} = \frac{100\cdot 99\cdot 98\cdot 97.96!}{4\cdot 3\cdot 2\cdot 1.96!} 3,921,225$$

21

# Organic Pollution-1

Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that, in the next 18 samples, exactly 2 contain the pollutant.

Let $X$ denote the number of samples that contain the pollutant in the next 18 samples analyzed. Then $X$ is a binomial random variable with $p = 0.1$ and $n = 18$

$$P(X = 2) = \binom{18}{2}(0.1)^2(0.9)^{16} = 153(0.1)^2(0.9)^{16} = 0.2835$$

| | |
|---|---|
| 0.2835 | = BINOMDIST(2,18,0.1,FALSE) |

22

## Organic Pollution-2

Determine the probability that at least 4 samples contain the pollutant.

$$P(X \geq 4) = 1 - P(X < 4)$$
$$= 1 - \sum_{x=0}^{3} \binom{18}{x} (0.1)^x (0.9)^{18-x}$$
$$= 1 - [0.150 + 0.300 + 0.284 + 0.168]$$
$$= 0.098$$

| | |
|---|---|
| 0.0982 | = 1 - BINOMDIST(3,18,0.1,TRUE) |



## Organic Pollution-3

Now determine the probability that $3 \leq X < 7$.

$$P(3 \leq X < 7) = \sum_{x=3}^{6} \binom{18}{x} (0.1)^x (0.9)^{18-x}$$
$$= 0.168 + 0.070 + 0.022 + 0.005$$
$$= 0.265$$

**0.265 = BINOMDIST(6,18,0.1,TRUE) - BINOMDIST(2,18,0.1,TRUE)**

## Binomial Mean and Variance

If $X$ is a binomial random variable with parameters $p$ and $n$,

$$\mu = E(X) = np$$

and

$$\sigma^2 = V(X) = np(1-p)$$



## Example:

For the coin toss example, $n = 3$ and $p = 0.8$, Find the mean and variance of the binomial random variable.

$\mu = E(X) = np = 3*0.8 = 2.4$ ⟹ Note that the expected value is exactly the same with the one found previously

$\sigma^2 = V(X) = np(1-p) = 3*0.2*0.8 = 0.48$

$\sigma = SD(X) = 0.693$

# Geometric Distribution

- Binomial distribution has:
  - Fixed number of trials.
  - Random number of successes.

- Geometric distribution has reversed roles:
  - Random number of trials.
  - Fixed number of successes, in this case 1.

- The probability density function of Geometric distribution is

$$f(x) = p(1-p)^{x-1}$$

$x = 1, 2, \ldots \infty$, the number of failures until the $1^{st}$ success. $0 < p < 1$, the probability of success.

27

# Wafer Contamination

The probability that a wafer contains a large particle of contamination is 0.01. Assume that the wafers are independent. What is the probability that exactly 125 wafers need to be analyzed before a particle is detected?

Let $X$ denote the number of samples analyzed until a large particle is detected. Then $X$ is a geometric random variable with parameter $p = 0.01$.

$$P(X=125) = (0.99)^{124}(0.01) = 0.00288.$$

28

## Geometric Mean & Variance

If $X$ is a geometric random variable with parameter $p$,

$$\mu = E(X) = \frac{1}{p}$$

and

$$\sigma^2 = V(X) = \frac{(1-p)}{p^2}$$



## Exercise: Mean and Standard Deviation

The probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume that the transmissions are independent events, and let the random variable $X$ denote the number of bits transmitted until the first error. Find the mean and standard deviation.

Mean = $\mu = E(X) = 1 / p = 1 / 0.1 = 10$

Variance = $\sigma^2 = V(X) = (1-p) / p^2 = 0.9 / 0.01 = 90$

Standard deviation = $\sqrt{90} = 9.49$

## Lack of Memory Property

For a geometric random variable, the trials are independent. Thus the count of the number of trials until the next success can be started at any trial without changing the probability distribution of the random variable.

The implication of using a geometric model is that the system presumably will not wear out. For all transmissions the probability of an error remains constant. Hence, the geometric distribution is said to lack any memory.

31

## Lack of Memory Property

The probability that a bit is transmitted in error is 0.1. Suppose 50 bits have been transmitted. What is the mean number of bits transmitted until the next error?

The mean number of bits transmitted until the next error, after 50 bits have already been transmitted, is

$$1 / 0.1 = 10.$$

the same result as the mean number of bits until the first error.

32

## Poisson Distribution

The random variable $X$ that equals the number of events in a Poisson process is a Poisson random variable with parameter λ > 0, and the probability density function is:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad \text{for} \quad x = 0, 1, 2, 3, \ldots$$

33

## Poisson Distribution

- Poisson Distribution: Expresses the probability of a given number of events occurring in a fixed interval of time and/or space.
- Events should occur with a known average rate and independently of the time since the last event.
- The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad \text{for} \quad x = 0, 1, 2, 3, \ldots$$

$\mu = E(X) = \lambda$ and $\sigma^2 = V(X) = \lambda$

34

## Calculations for Wire Flaws-1

For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per mm. Find the probability of exactly 2 flaws in 1 mm of wire.

Let *X* denote the number of flaws in 1 mm of wire

$$P(X=2)=\frac{e^{-2.3}\,2.3^2}{2!}=0.265$$

| In Excel | |
|---|---|
| 0.26518 | =POISSON(2, 2.3, FALSE) |

35

## Calculations for Wire Flaws-2

Determine the probability of 10 flaws in 5 mm of wire

Let *X* denote the number of flaws in 5 mm of wire.

$$E(X)=\lambda=5\text{ mm}\cdot\ 2.3\text{ flaws/mm}=11.5\text{ flaws}$$

$$\mathrm{P}(X=10)=e^{-11.5}\frac{11.5^{10}}{10!}=0.113$$

| In Excel | |
|---|---|
| 0.1129 | =POISSON(10, 11.5, FALSE) |

36

## Calculations for Wire Flaws-3

Determine the probability of at least 1 flaw in 2 mm of wire.

Let $X$ denote the number of flaws in 2 mm of wire.
Note that $P(X \geq 1)$ requires $\infty$ terms.

$$E(X) = \lambda = 2 \text{ mm} \cdot 2.3 \text{ flaws/mm} = 4.6 \text{ flaws}$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-4.6} \frac{4.6^0}{0!} = 0.9899$$

| In Excel | |
|---|---|
| 0.989948 | = 1 - POISSON(0, 4.6, FALSE) |



## Poisson Mean & Variance

If $X$ is a Poisson random variable with parameter $\lambda$, then

$$\mu = E(X) = \lambda \quad \text{and} \quad \sigma^2 = V(X) = \lambda$$

The mean and variance of the Poisson model are the same.

For example, if particle counts follow a Poisson distribution with a mean of 25 particles per square centimeter, the variance is also 25 and the standard deviation of the counts is 5 per square centimeter.

If the variance of a data is much greater (or less) than the mean, then the Poisson distribution would not be a good model for the distribution of the random variable.

## Poisson Example

If a number of accidents occurring on a highway each day is a Poisson random variable with $\lambda = 3$, what is the probability that no accidents occur today?

$P(X=0) = e^{-3}(3^0)/0! = e^{-3} = 0.05$

What is the probability that less than three accidents occur in the next two days?

$\lambda = 3*2 = 6$
$P(X<3) = P(X=0)+ P(X=1)+ P(X=2)$
$= e^{-6}(6^0)/0! + e^{-6}(6^1)/1! + e^{-6}(6^2)/2! = 0.06$



## Poisson Example

*Note:* Sometimes we have the length of interval to observe the events but not the number of events per unit interval. In such cases you have to pay attention to compute $\lambda$ correctly.

Suppose in a drive-through burger restaurant a car arrives, on the average, every 10 minutes. Further assume that cars arrive following a Poisson distribution. Find the probability that no car arrives for the next hour?

$\lambda = 6$ (not 10!)
$P(X=0) = e^{-6}(6^0)/0! = e^{-6} = 0.0025$

How to understand when to use which discrete distribution ?

- Key to Uniform Discrete Distribution: Outcomes have equal probability
- Key to Binomial and Geometric: We need to have an underlying Bernoulli trial (a random experiment with two outcomes)
  - Binomial distribution has fixed number of trials and random number of successes.
  - Geometric distribution has reversed roles: Random number of trials, fixed number of successes (just 1)
- Key to Poisson: Number of events in a given interval. If an analysis based on empirical data is to be conducted, make sure that the variance and mean values are not different than each other (more on this issue will be discussed later).

41

# Poisson Approximation to Binomial

Binomial distribution, in a sense, is a counting distribution as well. Can there be relation between binomial and Poisson distributions?

In fact, the answer is Yes.
For the Binomial distribution cases where $n$ is large and $p$ is small (why?), and np<10; we can use Poisson to approximate the underlying Binomial distribution using $\lambda = np$.

A certain genetic characteristic is seen in p=.002 percent of the population. In a sample of 1000 individuals, what is the probability that we observe 3 individuals possessing the characteristics?

np = 2
In Excel,

=BINOMDIST(3;1000;0,002;FALSE) = 0,1806

=POISSON(3;2;FALSE) = 0,1804

42