Random variables, expectation, and variance

DSE 210

Last lecture : Lecture 1

- Sets and Counting
  - Sets, Tuples
  - Set Operations - Union, Intersection
  - Permutations
  - Combinations
  - Hands on
  - Self - practice
- Probability Space
  - Sample space and probability of Outcomes
  - Examples
  - Hands on
  - Self - practice
- Multiple events, conditioning, and independence
  - Conditional Probability
  - Summation Rule
  - Bayes' rule
  - Independence
  - Hands on
  - Self - practice

Outline : Lecture 2

- **Random Variables, Expectation and Variance**
  - **Random Variable**
  - **Expected Value**
  - **Variance**
  - **Sampling**
  - **Hands on**
  - **Self - practice**
- Modeling data with Probability distributions
  - Binomial distribution
  - MLE
  - Normal Distribution
  - Multinomial Distribution
  - Poisson Distribution
  - Hands on
  - Self - practice
- Linear Algebra Primer
  - Vectors and Matrices
  - Quadratic Functions
  - Hands on
  - Self - practice

Random variables

Roll a die.

$$\text{Define } X = \begin{cases} 1 & \text{if die is } \geq 3 \\ 0 & \text{otherwise} \end{cases}$$

Here the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.

$$\omega = 1, 2 \Rightarrow X = 0$$
$$\omega = 3, 4, 5, 6 \Rightarrow X = 1$$

Roll n dice.

$$X = \#\text{ of 6's}$$
$$Y = \#\text{ of 1's before the first 6}$$

Both X and Y are defined on the same sample space, $\Omega = \{1, 2, 3, 4, 5, 6\}^n$. For instance,

$$\omega = (1, 1, 1, \ldots, 1, 6) \;\Rightarrow\; X = 1, Y = n - 1.$$

In general, a **random variable (r.v.)** is a defined on a probability space. It is a mapping from Ω to $\mathbb{R}$. We'll use capital letters for r.v.'s.

The distribution of a random variable

Roll a die. Define $X = 1$ if die is ≥ 3, otherwise $X = 0$.

X takes values in $\{0, 1\}$ and has distribution:

$$\Pr(X = 0) = \frac{1}{3} \text{ and } \Pr(X = 1) = \frac{2}{3}.$$

Roll n dice. Define $X =$ number of 6's.

X takes values in $\{0, 1, 2, \ldots, n\}$. The distribution of X is:

$$\begin{aligned}\Pr(X = k) &= \#(\text{sequences with } k \text{ 6's}) \cdot \Pr(\text{one such sequence}) \\ &= \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}\end{aligned}$$

Throw a dart at a dartboard of radius 1. Let X be the distance to the center of the board.

X takes values in $[0, 1]$. The distribution of X is:

$$\Pr(X \leq x) = x^2.$$

Expected value, or mean

The expected value of a random variable X is

$$\mathbb{E}(X) = \sum_x x \Pr(X = x).$$

Roll a die. Let X be the number observed.

$$\begin{aligned}\mathbb{E}(X) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} \\ &= \frac{1+2+3+4+5+6}{6} = 3.5 \qquad \text{(average)}\end{aligned}$$

Biased coin. A coin has heads probability p. Let X be 1 if heads, 0 if tails.

$$\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p.$$

Toss a coin with bias p repeatedly, until it comes up heads.
Let X be the number of tosses.

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

Pascal's wager

Pascal: I think there is some chance ($p > 0$) that God exists. Therefore I should act as if he exists.

Let $X =$ my level of suffering.

- Suppose I behave as if God exists (that is, I behave myself).
  Then X is some significant but finite amount, like 100 or 1000.
- Suppose I behave as if God doesn't exists (I do whatever I want to).
  If indeed God doesn't exist: $X = 0$.
  But if God exists: $X = \infty$ (hell).
  Therefore, $\mathbb{E}(X) = 0 \cdot (1 - p) + \infty \cdot p = \infty$.

The first option is much better!

Linearity of expectation

- If you double a set of numbers, how is the average affected?

  It is also doubled.

- If you increase a set of numbers by 1, how much does the average change?

  It also increases by 1.

- Rule: $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ for any random variable X and any constants a, b.

- But here's a more surprising (and very powerful) property: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ for any two random variables X, Y.

- Likewise: $\mathbb{E}(X + Y + Z) = \mathbb{E}(X) + \mathbb{E}(Y) + \mathbb{E}(Z)$, etc.

Linearity: examples

Roll 2 dice and let Z denote the sum. What is $\mathbb{E}(Z)$?

Method 1

Distribution of Z:

z	2	3	4	5	6	7	8	9	10	11	12
$\Pr(Z = z)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Now use formula for expected value:

$$\mathbb{E}(Z) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \cdots = 7.$$

Method 2

Let X_1 be the first die and X_2 the second die. Each of them is a single die and thus (as we saw earlier) has expected value 3.5. Since $Z = X_1 + X_2$,

$$\mathbb{E}(Z) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 3.5 + 3.5 = 7.$$

Toss n coins of bias p, and let X be the number of heads. What is $\mathbb{E}(X)$?

Let the individual coins be $X_1, \ldots, X_n$.
Each has value 0 or 1 and has expected value p.

Since $X = X_1 + X_2 + \cdots + X_n$,

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np.$$

Roll a die n times, and let X be the number of 6's. What is $\mathbb{E}(X)$?

Let X_1 be 1 if the first roll is a 6, and 0 otherwise.

$$\mathbb{E}(X_1) = \frac{1}{6}.$$

Likewise, define $X_2, X_3, \ldots, X_n$.

Since $X = X_1 + \cdots + X_n$, we have

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = \frac{n}{6}.$$

Coupon collector, again

Each cereal box has one of k action figures. What is the expected number of boxes you need to buy in order to collect all the figures?

Suppose you've already collected $i-1$ of the figures. Let X_i be the time to collect the next one.

Each box you buy will contain a new figure with probability $(k-(i-1))/k$. Therefore,

$$\mathbb{E}(X_i) = \frac{k}{k-i+1}.$$

Total number of boxes bought is $X = X_1 + X_2 + \cdots + X_k$, so

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_k) \\
&= \frac{k}{k} + \frac{k}{k-1} + \frac{k}{k-2} + \cdots + \frac{k}{1} \\
&= k\left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right) \quad \approx \quad k \ln k.
\end{aligned}$$

Independent random variables

Random variables X, Y are independent if
$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

Independent or not?

- Pick a card out of a standard deck. $X =$ suit and $Y =$ number.

  Independent.

- Flip a fair coin n times. $X = \#$ heads and $Y =$ last toss.

  Not independent.

- X, Y take values $\{-1, 0, 1\}$, with the following probabilities:

		Y		
		-1	0	1
X	-1	0.4	0.16	0.24
	0	0.05	0.02	0.03
	1	0.05	0.02	0.03

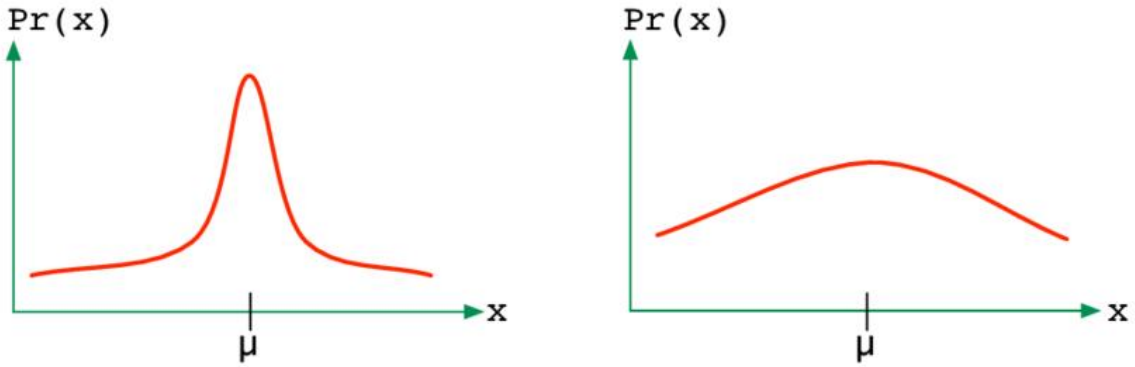
	X	Y
-1	0.8	0.5
0	0.1	0.2
1	0.1	0.3

Independent.

Variance

If you had to summarize the entire distribution of a r.v. X by a single number, you would use the mean (or median). Call it μ.

But these don't capture the *spread* of X:



What would be a good measure of spread? How about the average distance away from the mean: $\mathbb{E}(|X - \mu|)$?

For convenience, take the square instead of the absolute value.

Variance: $\quad \text{var}(X) = \mathbb{E}(X - \mu)^2 = \mathbb{E}(X^2) - \mu^2,$

where $\mu = \mathbb{E}(X)$. The variance is always ≥ 0.

Variance: example

Recall: $\text{var}(X) = \mathbb{E}(X-\mu)^2 = \mathbb{E}(X^2) - \mu^2$, where $\mu = \mathbb{E}(X)$.

Toss a coin of bias p. Let $X \in \{0, 1\}$ be the outcome.

$$\begin{aligned}
\mathbb{E}(X) &= p \\
\mathbb{E}(X^2) &= p \\
\mathbb{E}(X-\mu)^2 &= p^2 \cdot (1-p) + (1-p)^2 \cdot p = p(1-p) \\
\mathbb{E}(X^2) - \mu^2 &= p - p^2 = p(1-p)
\end{aligned}$$

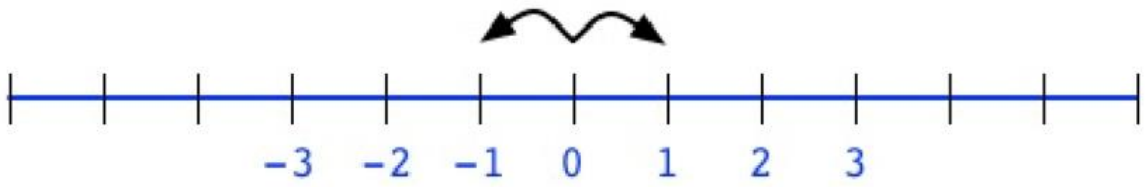
This variance is highest when $p = 1/2$ (fair coin).

The standard deviation of X is $\sqrt{\text{var}(X)}$.
It is the average amount by which X differs from its mean.

Variance of a sum

$\text{var}(X_1 + \cdots + X_k) = \text{var}(X_1) + \cdots + \text{var}(X_k)$ if the X_i are independent.

Symmetric random walk. A drunken man sets out from a bar. At each time step, he either moves one step to the right or one step to the left, with equal probabilities. Roughly where is he after n steps?



Let $X_i \in \{-1, 1\}$ be his ith step. Then $\mathbb{E}(X_i) = 0$ and $\text{var}(X_i) = 1$.

His position after n steps is $X = X_1 + \cdots + X_n$.

$$\mathbb{E}(X) = 0$$
$$\text{var}(X) = n$$
$$\text{stddev}(X) = \sqrt{n}$$

He is likely to be pretty close to where he started!

Sampling

Useful variance rules:

- $\mathrm{var}(X_1 + \cdots + X_k) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_k)$ if X_i's independent.
- $\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$.

What fraction of San Diegans like sushi? Call it p.

Pick n people at random and ask them. Each answers 1 (likes) or 0 (doesn't like). Call these values $X_1, \ldots, X_n$. Your estimate is then:

$$Y = \frac{X_1 + \cdots + X_n}{n}.$$

How accurate is this estimate?

Each X_i has mean p and variance $p(1-p)$, so

$$\mathbb{E}(Y) = \frac{\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)}{n} = p$$

$$\mathrm{var}(Y) = \frac{\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)}{n^2} = \frac{p(1-p)}{n}$$

$$\mathrm{stddev}(Y) = \sqrt{\frac{p(1-p)}{n}} \leq \frac{1}{2\sqrt{n}}$$

Outline : Lecture 2

- **Random Variables, Expectation and Variance**
  - **Random Variable**
  - **Expected Value**
  - **Variance**
  - **Sampling**
  - **Hands on**
  - **Self - practice**
- Modeling data with Probability distributions
  - Binomial distribution
  - MLE
  - Normal Distribution
  - Multinomial Distribution
  - Poisson Distribution
  - Hands on
  - Self - practice
- Linear Algebra Primer
  - Vectors and Matrices
  - Quadratic Functions
  - Hands on
  - Self - practice

Modeling data with probability distributions

DSE 210

Outline : Lecture 2

- Random Variables, Expectation and Variance
  - Random Variable
  - Expected Value
  - Variance
  - Sampling
  - Hands on
  - Self - practice
- **Modeling data with Probability distributions**
  - **Binomial distribution**
  - **MLE**
  - **Normal Distribution**
  - **Multinomial Distribution**
  - **Poisson Distribution**
  - **Hands on**
  - **Self - practice**
- Linear Algebra Primer
  - Vectors and Matrices
  - Quadratic Functions
  - Hands on
  - Self - practice

Distributional modeling

A useful way to summarize a data set:

- Fit a probability distribution to it.
- Simple and compact, and captures the big picture while smoothing out the wrinkles in the data.
- In subsequent application, use distribution as a proxy for the data.

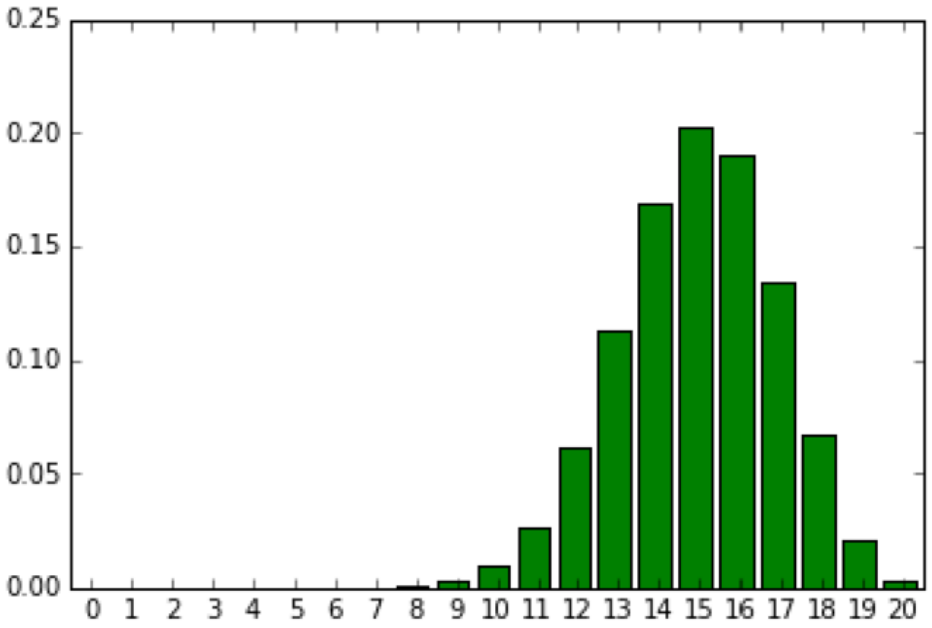
Which distributions to use?

> *There exist a few distributions of great universality which occur in a surprisingly large number of problems. The three principal distributions, with ramifications throughout probability theory, are the binomial distribution, the normal distribution, and the Poisson distribution.* – William Feller.

Well, this is true in one dimension. For higher-dimensional data, we'll use combinations of 1-d models: **products** and **mixtures**.

The binomial distribution

Binomial(n, p): the number of heads when n coins of bias (heads probability p) are tossed, independently.



Suppose X has a binomial(n, p) distribution.

$$\mathbb{E}X = np$$

$$\text{var}(X) = np(1-p)$$

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Fitting a binomial distribution to data

Example: Upcoming election in a two-party country.

- You choose 1000 people at random and poll them.
- 600 say Democratic.

What is a good estimate for the fraction of votes the Democrats will get in the election? Clearly, 60%.

More generally, you observe n tosses of a coin of unknown bias. k of them are heads. How to estimate the bias?

$$p = \frac{k}{n}$$

Geometric Distribution

Geometric $\mathrm{P}(X = n) = \mathrm{P}(\text{first success at } n\text{'th trial})$
Suppose X has a Geometric $\mathrm{P}(X = n)$ distribution.

$$\mathbb{E}X = \frac{1}{p}$$

$$\mathrm{var}(X) = \frac{1-p}{p^2}$$

$$\Pr(X = n) = p(1-p)^{n-1}$$

Geometric Distribution Example

Example: If the probability of a start-up being successful is P(start-up success) $= 20\%$, what is the expected # of start-ups until we observe the first success?

$X \sim \textit{Geometric}(0.2)$

$$E(X) = \frac{1}{0.2} = 5$$

Maximum likelihood estimation

Let $\mathcal{P}$ be a class of probability distributions (Gaussians, Poissons, etc).

Maximum likelihood principle: pick the distribution in $\mathcal{P}$ that makes the data maximally likely.
That is, pick the $p \in \mathcal{P}$ that maximizes $\Pr(\text{data}|p)$.

E.g. Suppose $\mathcal{P}$ is the class of binomials. We observe n coin tosses, and k of them are heads.

- Maximum likelihood : pick the bias p that maximizes

$$\Pr(\text{data}|p) = p^k(1-p)^{n-k}.$$

- Maximizing this is the same as maximizing its log,

$$\text{LL}(p) = k \ln p + (n-k)\ln(1-p).$$

- Set the derivative to zero.

$$\text{LL}'(p) = \frac{k}{p} - \frac{n-k}{1-p} = 0 \quad \Rightarrow \quad p = \frac{k}{n}.$$

Maximum likelihood: a small caveat

You have two coins of unknown bias.

- You toss the first coin 10 times, and it comes out heads every time. You estimate its bias as $p_1 = 1.0$.
- You toss the second coin 10 times, and it comes out heads once. You estimate its bias as $p_2 = 0.1$.

Now you are told that one of the coins was tossed 20 times and 19 of them came out heads. Which coin do you think it is?

- Likelihood under p_1:

$$\Pr(\text{19 heads out of 20 tosses} | \text{bias} = 1) = (1)^{19}(0)^1 = 0$$

- Likelihood under p_2:

$$\Pr(\text{19 heads out of 20 tosses} | \text{bias} = 0.1) = (0.1)^{19}(0.9)^1 = 9 \cdot 10^{-20}$$

The likelihood principle would choose the second coin. Is this right?

Laplace smoothing

A smoothed version of maximum-likelihood: when you toss a coin n times and observe k heads, estimate the bias as

$$p = \frac{k+1}{n+2}.$$

Laplace's law of succession: What is the probability that the sun won't rise tomorrow?

- Let p be the probability that the sun won't rise on a randomly chosen day. We want to estimate p.
- For the past 5000 years ($= 1825000$ days), the sun has risen every day. Using Laplace smoothing, estimate

$$p = \frac{1}{1825002}.$$

Coin Example Revisited

You have two coins of unknown bias.

- You toss the first coin 10 times, and it comes out heads every time.

  Using Laplace smoothing, you estimate its bias as $p_1 = \frac{11}{12}$.
- You toss the second coin 10 times, and it comes out heads once.

  Using Laplace smoothing, you estimate its bias as $p_2 = \frac{2}{12}$.

Now you are told that one of the coins was tossed 20 times and 19 of them came out heads. Which coin do you think it is?

- Likelihood under p_1:

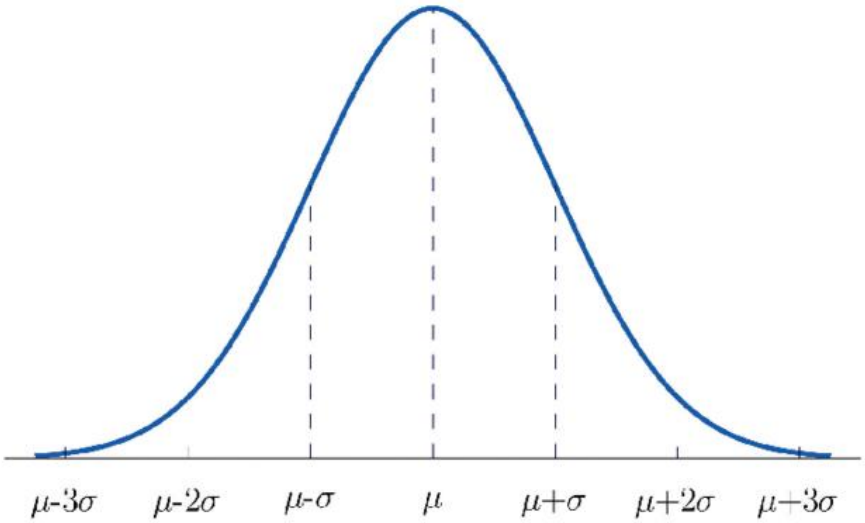
$$\Pr(\text{19 heads out of 20 tosses} | \text{bias} = 11/12) = \left(\frac{11}{12}\right)^{19} \left(\frac{1}{12}\right)^{1} \approx 0.016$$

- Likelihood under p_2:

$$\Pr(\text{19 heads out of 20 tosses} | \text{bias} = 2/12) = \left(\frac{2}{12}\right)^{19} \left(\frac{10}{12}\right)^{1} \approx 1.4 \cdot 10^{-15}$$

The answer changes if we use Laplace smoothing instead of MLE while estimating the biases.

The normal distribution



The normal (or *Gaussian*) $N(\mu, \sigma^2)$ has mean μ, variance σ^2, and density function

$$p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

- 68.3% of the distribution lies within one standard deviation of the mean, i.e. in the range $\mu \pm \sigma$
- 95.5% lies within $\mu \pm 2\sigma$
- 99.7% lies within $\mu \pm 3\sigma$

Maximum likelihood estimation of the normal

Suppose you see n data points $x_1, \ldots, x_n \in \mathbb{R}$, and you want to fit a Gaussian $N(\mu, \sigma^2)$ to them. How to choose μ, σ?

- Maximum likelihood: pick μ, σ to maximize

$$\Pr(\text{data}|\mu, \sigma^2) = \prod_{i=1}^{n} \left(\frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right)$$

  Note that this is found by multiplying probability density functions of n data points.

- Work with the log, since it makes things easier:

$$\text{LL}(\mu, \sigma^2) = \frac{n}{2} \ln \frac{1}{2\pi\sigma^2} - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}.$$

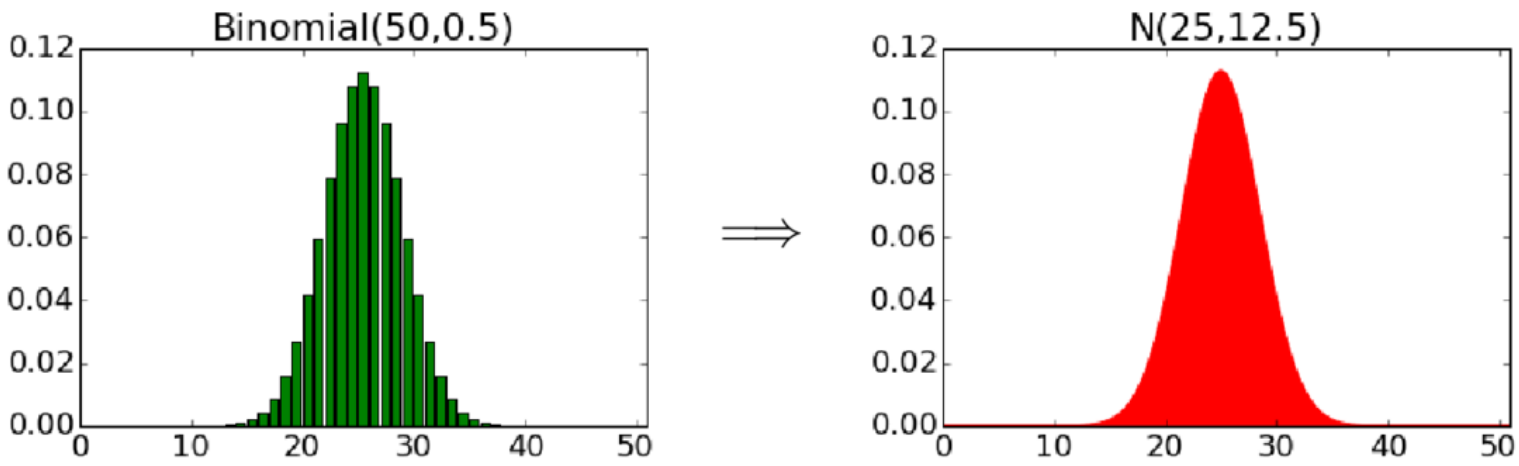
- Setting the derivatives to zero, we get

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

These are simply the empirical mean and variance.

Normal approximation to the binomial



When a coin of bias p is tossed n times, let X be the number of heads.

- We know X has mean np and variance $np(1-p)$.
- As n grows, the distribution of X looks increasingly like a Gaussian with this mean and variance.

Application to sampling

We want to find out what fraction p of San Diegans know how to surf. So we poll n random people, and find that k of them surf. Our estimate:

$$\widehat{p} = \frac{k}{n}.$$

Normal approximation:

- k has a binomial(n, p) distribution.
- This is close to a Gaussian with mean np and variance $np(1-p)$.
- Therefore the distribution of $\widehat{p} = k/n$ is close to a Gaussian with

$$\text{mean} = p$$

$$\text{variance} = \frac{p(1-p)}{n} \leq \frac{1}{4n}$$

Confidence intervals: Setting $\sigma = \sqrt{1/4n} = 1/(2\sqrt{n})$. Using 68, 95, 99 Rule:

- With 95% confidence, our estimate is accurate within $\pm 1/\sqrt{n}$.
- With 99% confidence, our estimate is accurate within $\pm 3/2\sqrt{n}$.

The multinomial distribution

A k-sided die:

- A fair coin has two possible outcomes, each equally likely.
- A fair die has six possible outcomes, each equally likely.
- Imagine a k-faced die, with probabilities $p_1, \ldots, p_k$.

Toss such a die n times, and count the number of times each of the k faces occurs:

$$X_j = \#\text{ of times face } j \text{ occurs}$$

The distribution of $X = (X_1, \ldots, X_k)$ is called the **multinomial**.

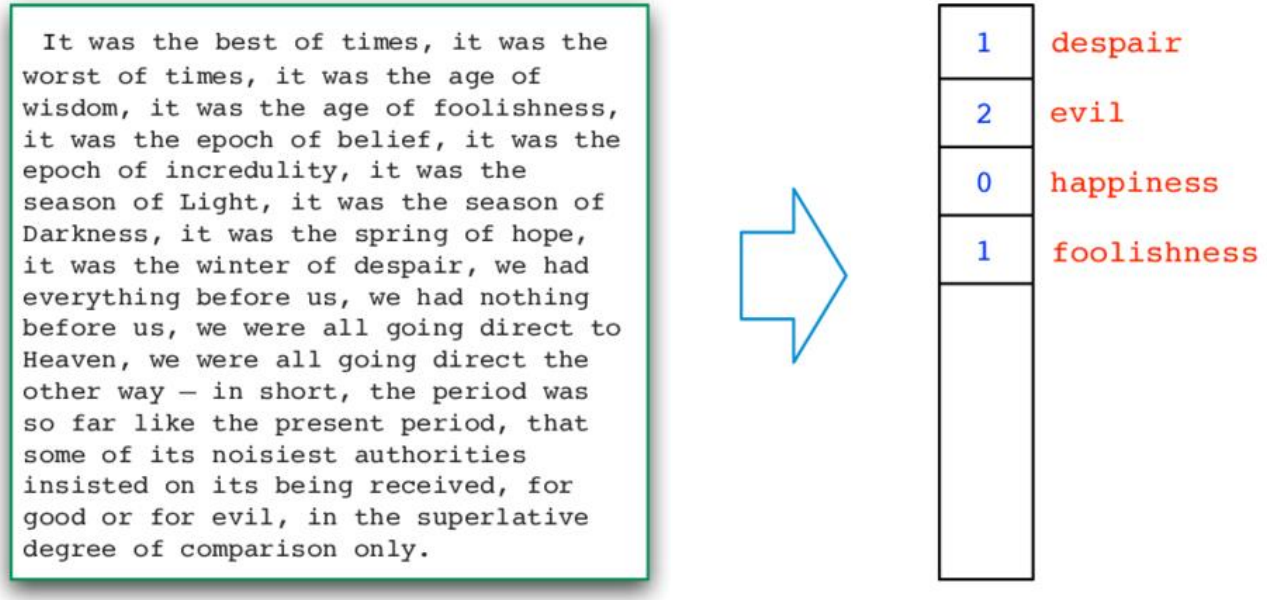
- Parameters: $p_1, \ldots, p_k \geq 0$, with $p_1 + \cdots + p_k = 1$.
- $\mathbb{E}X = (np_1, np_2, \ldots, np_k)$.
- $\Pr(n_1, \ldots, n_k) = \binom{n}{n_1, n_2, \ldots, n_k} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$, where

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}$$

is the number of ways to place balls numbered $\{1, \ldots, n\}$ into bins numbered $\{1, \ldots, k\}$.

Example: text documents

Bag-of-words: vectorial representation of text documents.

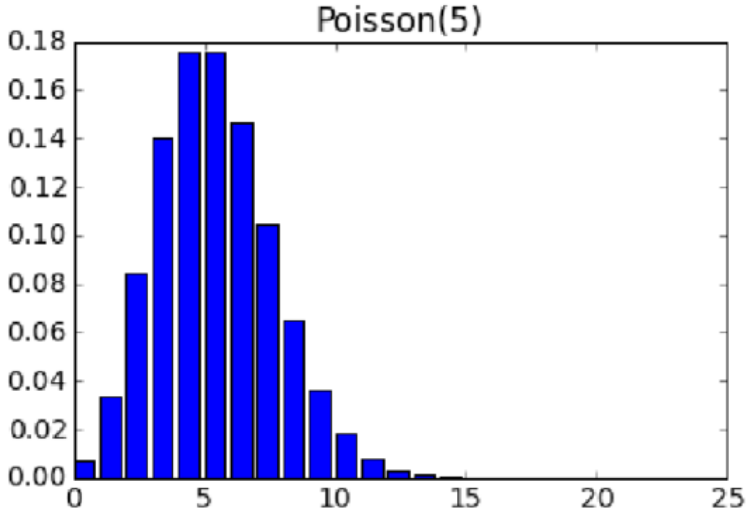


- Fix V = some vocabulary.
- Treat the words in a document as independent draws from a multinomial distribution over V:

$$p = (p_1, \ldots, p_{|V|}), \quad \text{such that} \quad p_i \geq 0 \text{ and } \sum_i p_i = 1$$

The Poisson distribution

A distribution over the non-negative integers $\{0, 1, 2, \ldots\}$



The Poisson has parameter $\lambda > 0$, with $\Pr(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$

- Mean: $\mathbb{E}X = \lambda$
- Variance: $\mathbb{E}(X - \lambda)^2 = \lambda$
- Maximum likelihood fit: set λ to the empirical mean

How the Poisson arises

Count the number of events (collisions, phone calls, etc) that occur in a certain interval of time. Call this number X, and say it has expected value λ.

Now suppose we divide the interval into small pieces of equal length.

If the probability of an event occurring in a small interval is:

- independent of what happens in other small intervals, and
- the same across small intervals,

then $X \sim \text{Poisson}(\lambda)$.

Deriving Poisson Formula

P_λ approximates $B_{p,n}$ for $\lambda = pn$, when $n \gg 1 \gg p$.

$B_{p,n}(k) = \binom{n}{k} p^k q^{n-k}$, where $p = \frac{\lambda}{n}$ and $q = 1 - p$

$$= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n(n-1)...(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^k}$$

Deriving Poisson Formula

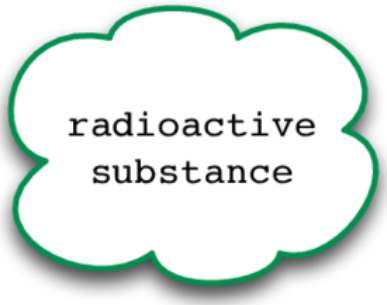
Limit of Binomial, λ and k fixed, as n $\longrightarrow \infty$:

$$B_{p,n}(k) = \frac{n(n-1)\ldots(n-k+1)}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \longrightarrow e^{-\lambda}\frac{\lambda^k}{k!}$$

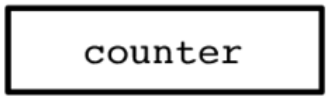
- $\frac{(n)(n-1)\ldots(n-k+1)}{(n)(n)\ldots(n)} \longrightarrow 1$ (fixed # of k terms, each 1)
- $\left(1-\frac{\lambda}{n}\right)^k \longrightarrow 1$ (fixed # of k terms, each 1)
- $\left(1-\frac{\lambda}{n}\right)^n = \left(\left(1-\frac{\lambda}{n}\right)^{n/\lambda}\right)^{\lambda} = (e^{-1})^{\lambda} = e^{-\lambda}$

Poisson: examples

Rutherford's experiments with radioactive disintegration (1920)



- $N = 2608$ intervals of 7.5 seconds
- $N_k = \#$ intervals with k particles
- Mean: 3.87 particles per interval



k	0	1	2	3	4	5	6	7	8	≥ 9
N_k	57	203	383	525	532	408	273	139	45	43
P(3.87)	54.4	211	407	526	508	394	254	140	67.9	46.3

Flying bomb hits on London in WWII

- Area divided into 576 regions, each 0.25 km^2
- $N_k = \#$ regions with k hits
- Mean: 0.93 hits per region

k	0	1	2	3	4	≥ 5
N_k	229	211	93	35	7	1
P(0.93)	226.8	211.4	98.54	30.62	7.14	1.57

Multivariate distributions

Almost all distributions we've considered are for one-dimensional data.

- Binomial, Poisson: integer
- Gaussian: real

What to do with the usual situation of data in higher dimensions?

1. Model each coordinate separately and treat them as independent.
   For $x = (x_1, \ldots, x_p)$, fit separate models $\Pr_i$ to each x_i, and assume

$$\Pr(x_1, \ldots, x_p) = \Pr_1(x_1)\Pr_2(x_2)\cdots\Pr_p(x_p).$$

   This assumption is almost always completely inaccurate, and sometimes causes problems.

2. Multivariate Gaussian.
   Allows modeling of correlations between coordinates.

3. More general graphical models.
   Arbitrary dependencies between coordinates.

Multivariate distributions: Multivariate Gaussian

Also known as Multivariate Normal Distribution.

Generalization of the one-dimensional (uni-variate) normal distribution to higher dimensions.

$$
\begin{aligned}
f(x_i|\mu,\Sigma) &= \\
&= f((x_{i1}, x_{i2}, ..., x_{in})' \,|\, \mu = (\mu_1, ..., \mu_n)', \Sigma = \begin{bmatrix} \sigma_1^2 & & \sigma_{1n} \\ & \ddots & \\ \sigma_{n1} & & \sigma_n^2 \end{bmatrix}) \\
&= \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} exp(-\frac{1}{2}(x_i - \mu)' \Sigma^{-1} (x_i - \mu))
\end{aligned}
$$

We will revisit Multivariate Gaussian in the next lectures.

Appendix: MLE of Binomial Distribution

Suppose that X is an observation from a binomial distribution, X $\sim$ Binomial(n,p), where n is known and p is to be estimated. The likelihood function is:

- L(p)= $\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$

  Example: Likelihood of observing k heads when a biased coin (p) tossed n times.

- Since the likelihood function is regarded as a function only of the parameter p, the factor $\frac{n!}{k!(n-k)!}$ is a fixed constant and does not affect the MLE.
- L(p)= $p^k(1-p)^{n-k}$

Appendix: MLE of Geometric Distribution

- Let $X_1, X_2, X_3.....X_n$ be a random sample from the geometric distribution with p.d.f.
  $f(x;p) = (1-p)^{x-1}p, x = 1, 2, 3....$
- The likelihood function is given by:
  $L(p) = (1-p)^{x_1-1}p(1-p)^{x_2-1}p...(1-p)^{x_n-1}p = p^n(1-p)^{\sum_1^n x_i - n}$
- Taking log,
  $lnL(p) = nlnp + \left(\sum_1^n x_i - n\right) ln(1-p)$
- Differentiating and equating to zero, we get,
  $\frac{d[lnL(p)]}{dp} = \frac{n}{p} - \frac{\left(\sum_1^n x_i - n\right)}{(1-p)} = 0$
- Therefore,
  $p = \frac{n}{\left(\sum_1^n x_i\right)}$
- So, the maximum likelihood estimator of P is:
  $P = \frac{n}{\left(\sum_1^n X_i\right)} = \frac{1}{X}$
- This agrees with the intuition because, in n observations of a geometric random variable, there are n successes in the $\sum_1^n X_i$ trials. Thus the estimate of p is the number of successes divided by the total number of trials.

Appendix: MLE of Poisson Distribution

- Suppose that $X = (X_1, X_2, ..., X_n)$ are iid observations from a Poisson distribution with unknown parameter λ. The likelihood function is:

$$\begin{aligned} L(\lambda; x) &= \prod_{i=1}^{n} f(x_i; \lambda) \\ &= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \\ &= \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{x_1! x_2! \cdots x_n!} \end{aligned}$$

- By differentiating the log of this function with respect to λ, that is by differentiating the Poisson log likelihood function

  $l(\lambda; x) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda$

  ignoring the constant terms that do not depend on λ, one can show that the maximum is achieved at $\hat{\lambda} = \sum_{i=1}^{n} x_i/n$. Thus, for a Poisson sample, the MLE for λ is just the sample mean.

Outline : Lecture 2

- Random Variables, Expectation and Variance
  - Random Variable
  - Expected Value
  - Variance
  - Sampling
  - Hands on
  - Self - practice
- **Modeling data with Probability distributions**
  - **Binomial distribution**
  - **MLE**
  - **Normal Distribution**
  - **Multinomial Distribution**
  - **Poisson Distribution**
  - **Hands on**
  - **Self - practice**
- Linear Algebra Primer
  - Vectors and Matrices
  - Quadratic Functions
  - Hands on
  - Self - practice