Victory Lap!

CSE 312 24su

"when life gives you lemons...use CLT to find how many lemons you need to make a good lemonade"

Logistics

- Midterm “makeup” clarification
- Final exam tomorrow
- Claris’s office hours tomorrow will start at 12:30 (not 12pm)
- Course evals reopened – fill out tonight for +1 point
- See Ed post

One day...there was *counting*

We learned about tecniques to count the number of possible outcomes in a set

Lots of counting...

Factorial: $n!$ ways to rearrange n distinct things

Complementary counting: Counting the ways for A to **not** occur
$\text{ways for A to NOT occur} = \text{total options} - \text{ways for A to occur}$

Product rule: *Sequential process* with m_1 options in 1st step, m_2 options in 2nd , m_3 in 3rd step, etc. we pick 1 option from each to form the outcome $m_1 \cdot m_2 \cdot m_3 \cdot \ldots$

Stars and bars: $\binom{n+k-1}{k-1}$ ways to distribute n identical things to k distinct types

Picking k distinct elements from a group of n distinct elements

Permutations: $P(n,k)$ if the **order of the k elements does matter**

Combinations: $\binom{n}{k}$ if the order of the k elements does **not** matter

Finding the size of a union of sets - $|A \cup B \cup \cdots|$

Sum rule: If disjoint, $|A| + |B| + \cdots$

Inclusion-Exclusion: *singles-doubles+triples-...*

Counting Leads to...Probability

A **(discrete) probability space** is a pair $(\Omega, \mathbb{P})$ where:

- $\boldsymbol{\Omega}$ is the sample space
- $\mathbb{P}: \Omega \to [0,1]$ is the probability measure.

Uniform probability space: *Every outcome equally likely to occur*

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|}, \; \mathbb{P}(E) = \frac{|E|}{|\Omega|}$$

> Pick a sample space where every outcome is **equally likely**
> Find the size of the sample space (using counting techniques!)
> Define the event and count its size (using counting techniques!)
> Find the probability by doing $\mathbb{P}(E) = \frac{|E|}{|\Omega|}$

When we're given extra info…

Conditional Probability – "restrict the sample space"

Definition of cond. prob.: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Bayes' theorem: $\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(B)}$

Law of total probability:

$\mathbb{P}(A) = \mathbb{P}(A|E_1)\mathbb{P}(E_1) + \cdots + \mathbb{P}(A|E_n)\mathbb{P}(E_n)$
if $E_1, E_2, \ldots, E_n$ partition the sample space Ω

Chain Rule: $\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) =$
$\mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(\mathrm{E}_3|\mathrm{E}_1 \cap E_2) \ldots$

maybe that info was irrelevant...

Independence

If events A and B are *<u>independent</u>*

$$\mathbb{P}(A|B) = \mathbb{P}(B), \mathbb{P}(\mathrm{A} \cap B)\mathbb{P}(A)\mathbb{P}(B)$$

If events $A, B, C, \ldots$ are *<u>mutually independent</u>*

$$\mathbb{P}(\mathrm{A} \cap B \cap C \cap \cdots) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \ldots$$

If events A and B are *<u>conditionally independent</u>* on C:

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$$

Sometimes we're interested in analyzing quantitative properties

Random Variables

Outcome $\omega \in \Omega$ → X → Quantitative value

Range/Support

$\mathbf{\Omega_X}$ is set of possible values X can be

Probability Mass Function

$$p_X(k) = \mathbb{P}(X = k)$$

Cumulative Distribution Function

$$F_X(k) = \mathbb{P}(X \leq k)$$

Expectation

Weighted average of values in the support

Variance

Measure the spread of the distribution

maybe those quantitative properties are continuous...

Random Variables

Outcome $\omega \in \Omega$ → X → *Quantitative value*

Range/Support

$\mathbf{\Omega_X}$ is set of possible values X can be

Probability <u>Density</u> Function

$$f_X(k) = \mathbb{P}(X = k) = \frac{d}{dk} F_X(k)$$

Cumulative Distribution Function

$$F_X(k) = \mathbb{P}(X \leq k) = \int_{-\infty}^{k} f_X(x)\, dx$$

Expectation

Weighted average of values in the support

Variance

Measure the spread of the distribution

Computing *Expectations*

Definition of Expectation

Definition of Expectation

$\mathbf{\Omega_X}$ is set of possible values X can be

Linearity of Expectation

$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$

From the zoo of the random variables...

Law of the Unconscious Statistician

$\mathbb{E}[g(X)] = \sum_{k \in \Omega_X}(g(k) \cdot p_X(k))$, $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(z) \cdot f_X(z)\, \mathrm{d}z$

Law of Total Expectation

$\mathbb{E}[X] = \sum E[X|A_i]\mathbb{P}(A_i)$, $\sum_{k \in \Omega_X}(y \cdot \mathbb{E}[X|Y = y])$, $\int_{-\infty}^{\infty} y \cdot \mathbb{E}[X|Y = y]\mathrm{d}y$

Computing *Variance*

Definition of Variance:

$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Properties of Variance:

$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ *if* X and Y are independent

Finding all that can be tedious...

Zoo of *Discrete* Random Variables

$X \sim \text{Unif}(a, b)$

$$p_X(k) = \frac{1}{b-a+1}$$
$$\mathbb{E}[X] = \frac{a+b}{2}$$
$$\text{Var}(X) = \frac{(b-a)(b-a+2)}{12}$$

$X \sim \text{Ber}(p)$

$$p_X(0) = 1-p; \quad p_X(1) = p$$
$$\mathbb{E}[X] = p$$
$$\text{Var}(X) = p(1-p)$$

$X \sim \text{Bin}(n, p)$

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$\mathbb{E}[X] = np$$
$$\text{Var}(X) = np(1-p)$$

$X \sim \text{Poi}(\lambda)$

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$
$$\mathbb{E}[X] = \lambda$$
$$\text{Var}(X) = \lambda$$

$X \sim \text{Geo}(p)$

$$p_X(k) = (1-p)^{k-1} p$$
$$\mathbb{E}[X] = \frac{1}{p}$$
$$\text{Var}(X) = \frac{1-p}{p^2}$$

$X \sim \text{NegBin}(r, p)$

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$
$$\mathbb{E}[X] = \frac{r}{p}$$
$$\text{Var}(X) = \frac{r(1-p)}{p^2}$$

$X \sim \text{HypGeo}(N, K, n)$

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$
$$\mathbb{E}[X] = n\frac{K}{N}$$
$$\text{Var}(X) = \frac{K(N-K)(N-n)}{N^2(N-1)}$$

Finding all that can be tedious...

Zoo of *Continuous* Random Variables

$X \sim \text{Unif}(a, b)$	$X \sim \text{Exp}(\lambda)$	$X \sim \mathcal{N}(\mu, \sigma^2)$
$f_X(k) = \frac{1}{b-a}$ for $a \le k \le b$ $F_X(k) = \frac{x-a}{b-a}$ if $a \le k < b$ $\mathbb{E}[X] = \frac{a+b}{2}$ $\text{Var}(X) = \frac{(b-a)^2}{12}$	$f_X(k) = \lambda e^{-\lambda k}$ for $k \ge 0$ $F_X(k) = 1 - e^{-\lambda k}$ if $k \ge 0$ $\mathbb{E}[X] = \frac{1}{\lambda}$ $\text{Var}(X) = \frac{1}{\lambda^2}$	$f_X(k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $F_X(k) = \Phi\left(\frac{k-\mu}{\sigma}\right)$ $\mathbb{E}[X] = \mu$ $\text{Var}(X) = \sigma^2$

One RV is always in our face…

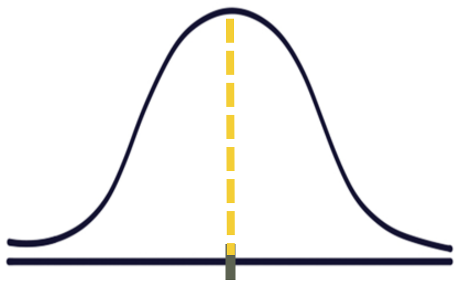
Normal Random Variables!

To compute probabilities with normal RVs:

1. **Standardize** the normal random variable: $Z = \frac{X-\mu}{\sigma}$
2. **Write probability expression in terms of $\Phi(\mathbf{z}) = \mathbb{P}(Z \leq z)$**
3. **Look up the value(s)** in the table

Central Limit Theorem

The sum of a bunch of i.i.d random variables can be **approximated** as a normal random variable.



More than one RV? We got it!

Joint Distributions

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)dsdt$
Normalization	$\sum_x \sum_y p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_y p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
Expectation	$E[g(X,Y)] = \sum_x \sum_y g(x,y)p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy$
Conditional PMF/PDF	$p_{X\mid Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$E[X \mid Y = y] = \sum_x x p_{X\mid Y}(x \mid y)$	$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X\mid Y}(x \mid y)dx$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x)p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x)f_Y(y)$

...and analyzing relationships

Covariance

$$\mathbf{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])] = \mathbb{E}[\boldsymbol{XY}] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{Y}]$$

Sometimes we know *nothing*...

(or we don't care) #relatable

Tail Bounds

- **Markov's inequality** - $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$
  - Use if X is non-negative and we know the expectation
- **Chebyshev's inequality** - $\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var(t)}}{t^2}$
  - Use if we know the expectation **and** variance of X
  - Gives better bounds with small variances
- **Chernoff Bound** $\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{\left(-\frac{\delta^2\mu}{2}\right)}$ *and* $\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{\left(-\frac{\delta^2\mu}{3}\right)}$
  - Use if X is a sum of independent Bernoulli random variables
- **Union Bound -** $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
  - Use if we don't have enough information to find the

Sometimes we know *nothing*...
(and we *want* to find out)

Maximum Likelihood Estimation

1. Write the likelihood function: $\mathcal{L}(x_1, \ldots, x_n; \theta)$
2. Take the log $ln(..)$ of the likelihood function
3. Take the derivative(s) of the log-likelihood function
4. Set the derivatives to 0, and solve for the MLE(s) $\hat{\theta}$
5. Verify it is a maximum with second derivative test
   (not required for 312)

Lots of applications!

- Naïve Bayes Spam Filtering
- Bloom Filters
- Efficient Distinct Elements
- Polling
- Multi-armed bandits (reinforcement learning)
- Randomized algorithm analysis
- Differential privacy

- Python and LaTeX

What's next?

I'm *really* going to miss this content...

- CSE **421**: Algorithms
  (design algos – uses some combinatorics)
- CSE **422**: Toolkit for Modern Algorithms
- CSE **426**: Cryptography
- CSE **427**: Computational Biology
- CSE **431**: Introduction to Theory of Computation
- CSE **446**: Machine Learning (probability + linear algebra)
- CSE **447**: Natural Language Processing
- CSE **473**: Artificial Intelligence (Bayes nets and such)
- CSE **490Q**: Quantum Computing

How do you want to spend class time today?

pollev.com/312

Maximum Likelihood Estimation

Suppose a wildlife researcher is studying the population of a rare bird species in a national park. The number of birds spotted each day follows a discrete probability mass function: $p_X(x;\theta) = \theta^x(1-\theta)$. The researcher observed the counts of birds for 30 days: $x_1, x_2, \ldots, x_{30}$. Find the MLE for θ.

1. Likelihood Function: $L(x_1,\ldots,x_{30};\theta) = \prod_{i=1}^{30}\left(\theta^{x_i}(1-\theta)\right)$

2. Log-likelihood: $\ln(L(\ldots)) = \sum_{i=1}^{30}\left(x_i \cdot \ln(\theta) + \ln(1-\theta)\right)$

3. Derivative: $\frac{d}{d\theta}(\ln(L(\ldots))) = \sum_{i=1}^{30} \frac{x_i}{\theta} - \frac{1}{1-\theta}$

4. Set to 0 & solve:

$$\sum_{i=1}^{30}\left(\frac{x_i}{\hat{\theta}} - \frac{1}{1-\hat{\theta}}\right) = 0 \Rightarrow \frac{\sum_{i=1}^{30} x_i}{\hat{\theta}} = \frac{30}{1-\hat{\theta}} \Rightarrow \sum_{i=1}^{30} x_i - \hat{\theta}\sum_{i=1}^{30} x_i = 30 \cdot \hat{\theta}$$

on the final, this is a valid final answer!

$$\Sigma = (30+\Sigma)\,\hat{\theta} \qquad \hat{\theta} = \frac{\Sigma}{30+\Sigma}$$

How to write the likelihood function?

"Given observations $x_1, x_2, \ldots, x_n$ from probability distribution X..."

we find the likelihood by **multiplying** together the probabilities or densities of each observation.

probabilities — if discrete; densities — if continuous

below are some examples:

e.g., if you're given the PMF/PDF like this, with only one case, and it's in terms of x_i,...

$p_x(k)$ (or $f_x(k)$) = [some exp. in terms of k] if $k \in \Omega_x$,

we see $x_1, \ldots, x_n$:

$$L(x_1, \ldots, x_n) = p_x(x_1;\theta) \times p_x(x_2;\theta) \times \ldots \times p_x(x_n;\theta)$$

$$= \prod_{i=1}^{n} p_x(x_i;\theta)$$

e.g., if you're given the PMF that has multiple cases, and isn't in terms of x_i

$$p_x(k) = \begin{cases} \theta_1 & k=1 \\ \theta_2 & k=2 \\ 1-\theta_1-\theta_2 & k=3 \end{cases}$$

← here, our PMF/PDF is NOT in terms of x_i, so we're going to be able to simplify the likelihood to NOT have that $\prod$

we see $x_1, \ldots, x_n$, 10 of which are 1, 20 of which are 2, and 15 of which are 3

$$L(x_1, \ldots, x_n) = p_x(x_1;\theta) \cdot p_x(x_2;\theta) \cdot \ldots \cdot p_x(x_n;\theta)$$

because of the counts given, we know 10 of these terms are θ_1, 20 of these terms are θ_2, and 15 are $1-\theta_1-\theta_2$. so....

$$= (\theta_1)^{10}(\theta_2)^{20}(1-\theta_1-\theta_2)^{15}$$

More Busy Servers

In a large server farm, 1800 requests are processed independently by a web server. The time taken for each request to be processed is an exponential random with parameter 0.5 seconds. → for each request, average of $\frac{1}{0.5} = 2$ sec.

We have HIGH expectations…a ***slow request*** is processed in more than 0.8 seconds.

Let X ~ number of slow requests

What is the expected number of slow requests? LOE!

1. Decompose: $X_i = \begin{cases} 1 & \text{if } i^{th} \text{ request is slow} \\ 0 & \text{otherwise} \end{cases}$ → $X = \sum_{i=1}^{1800} X_i$
2. Apply LOE $E[X] = E[\sum_{i=1}^{100} X_i] = \sum_{i=1}^{100} E[X_i]$
3. Conquer: $E[X_i] = P(X_i = 1) = P(R_i > 0.8) = 1 - P(R_i < 0.8) = 1 - (1 - e^{-0.5 \cdot 0.8}) = e^{-0.4}$

$E[X] = \sum_{i=1}^{1800} E[X_i] = 1800 \cdot e^{-0.4}$

What is the variance of the number of slow requests?

We know X is sum of independent X_i's, so….

$Var(X) = \sum_{i=1}^{1800} Var(X_i) = \sum_{i=1}^{1800} e^{-0.4}(1 - e^{-0.4}) = 1800(e^{-0.4})(1 - e^{-0.4})$

$X_i \sim Ber(e^{-0.4})$

More Busy Servers

In a large server farm, 1800 requests are processed independently by a web server. The time taken for each request to be processed is an exponential random with parameter 0.5 seconds.

We have low expectations…a ***slow request*** is processed in more than 0.8 seconds.

Let $X \sim$ number of slow request. $E[X] = \mu$, $Var(X) = \sigma^2$

Bound the probability there are at least 1500 slow requests using:

> Chebyshev's

$$P(X \geq 1500) = P(X - \mu \geq 1500 - \mu) \leq P(|X - \mu| \geq 1500 - \mu)$$

$$\leq \frac{\sigma^2}{(1500 - \mu)^2}$$

> Chernoff Bound $X = \sum X_i$, where the X_i's are independent. So….

$$P(X \geq 1500) \leq e^{\left(\frac{-\delta^2 \mu}{3}\right)}$$

$$1500 = (1 + \delta)\mu \rightarrow \delta = \frac{1500}{\mu} - 1$$

More Busy Servers

In a large server farm, 1800 requests are processed independently by a web server. The time taken for each request to be processed is an exponential random with parameter 0.5 seconds.

We have low expectations...a ***slow request*** is processed in more than 0.8 seconds.

X ~ number of slow requests. Let $\mu = E[X]$, $\sigma^2 = Var(X)$ from prev. parts.

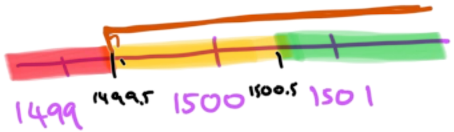
Approximate **the probability there are at least 1500 slow requests using:**

> Central Limit Theorem

1. Setup Problem: $X = \sum_{i=1}^{1800} X_i$, where $X_i \sim Ber(e^{-0.04})$ and are i.i.d.

We want $P(X \geq 1500)$ APPLY CONTINUITY CORRECTION

$= P(X \geq 1499.5)$



2. Apply CLT: X is sum of i.i.d, so by CLT, $X \approx Y \sim N(\mu, \sigma^2)$

So, $P(X \geq 1499.5) \approx P(Y \geq 1499.5)$

3. Solve. STANDARDIZE: $P(Y \geq 1499.5) = P\left(Z \geq \frac{1499.5 - \mu}{\sigma}\right)$

WRITE IN Φ $= 1 - \Phi\left(\frac{1499.5 - \mu}{\sigma}\right)$

SOLVE

Thank you all for an amazing quarter!! <3

have a great rest of the summer ☺