Review of Basic Probability Concepts (Chapters 1-4)

The goal of this course is to learn about *statistical inference*. Why do we need probability?

Statistical inference: learn about a population based on the information provided by a sample.

Population:

- Characterized by the probability distribution of a *random variable* X. Think of X as the value associated with a randomly selected individual from the population.
- The distribution of X depends on some unknown parameter θ.
- Goal: learn about θ.

Ex: What proportion of American voters think President Obama is doing a good job? Describe the population random variable and the parameter of interest.

Population = {All American voters}

Population: $X \sim \text{Bernoulli}(\theta)$

parameter of interest → θ = proportion of 1's in the population

sample (random)

$X_1, X_2, \ldots, X_n$

Define: $X = \begin{cases} 1, & \text{if President is doing a good job} \\ 0, & \text{otherwise} \end{cases} \sim \text{Bernoulli}(\theta)$

characteristic of interest for a (typical) randomly selected individual from this population

$\theta = P[X = 1]$

A general framework for inference

Population : $X \sim f_\theta(x)$

θ = parameter (may be a vector)

Goal: Learn about θ using information provided by the sample.

random sample of size n

$X_1, X_2, \ldots, X_n$

Two basic types of random variables

Discrete random variable: A rv X is *discrete* if the set of all possible values of X is countable.

Ex:

- X = indicator of heads in a coin toss ~ Bernoulli(θ)
  Possible values: $\{0,1\}$.
- X = rating of a professor on a scale of 1 to 5
- X = # people served in a drive-thru b/w 9 am – 9 pm.
  possible values: $\{0,1,2,\ldots\}$ – infinite set – countable.

Continuous random variable: A rv X is *continuous* if the set of all possible values of X is uncountable. — "interval"

Ex:

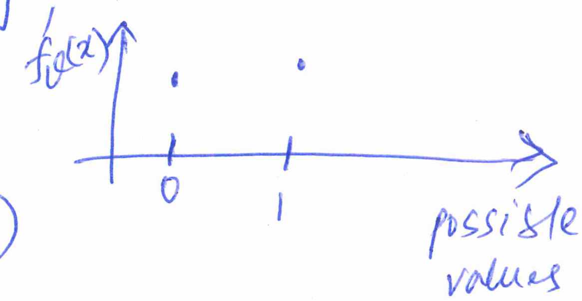
- X = lowest temp in Dallas on a day. Possible values: $(-\infty, \infty)$. ← uncountable set.
- X = Amplitude of sound waves: Possible values: $(0,\infty)$.
- X = service time for a customer
  Possible values: $(0,\infty)$

Discrete X

$\sum_x f_\theta(x) = 1 =$ total prob.

Probability distribution: $f_\theta(x) = P[X = x]$ $x \in \mathbb{R}$

- Probability mass function (PMF)



Computing probability: $P(X \in A) = \sum_{x \in A} f_\theta(x)$

Interpretation of $P(E)$, probability of an event E

- Long term proportion — if the experiment is repeated a large # times, then the proportion of times E will ~~occur~~ occur is approx. $P(E)$.

Expected value: $E(X) = \sum_x x f_\theta(x)$

Interpretation of $E(X)$

- long run average of X
- best guess of the value of X before the experiment
- Center of prob. dist. of X

Discrete X (cont'd)

Variance: $\mathrm{var}(X) =$ $E(X-\mu)^2$, $\mu = E(X)$.

Standard deviation: $\mathrm{sd}(X) =$ $\sqrt{\mathrm{var}(X)}$.

Interpretation of $\mathrm{var}(X)$ and $\mathrm{sd}(\cancel{x})$

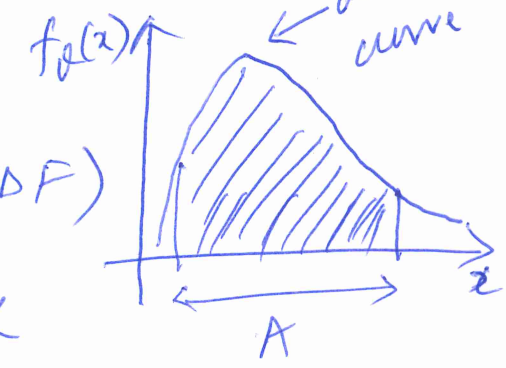
- ~~var~~ Measures of spread or uncertainty or variability in the dist. around the mean.
- $sd(X)$ tells how much off $E(X)$ is going to be from the true X.

Continuous X

$0 = P[X=x] \neq f_\theta(x)$

Probability distribution: $f_\theta(x)$, $\cancel{=}$ $x \in \mathbb{R}$

probability density function (PDF)



Computing probability: $P(X \in A) = \int_{x \in A} f_\theta(x)\,dx$

Expected value: $E(X) = \int_{-\infty}^{\infty} x f_\theta(x)\,dx$

Variance: $\text{var}(X)$ $\cancel{=}$ same as before.

<u>Interpretation:</u> same as before.

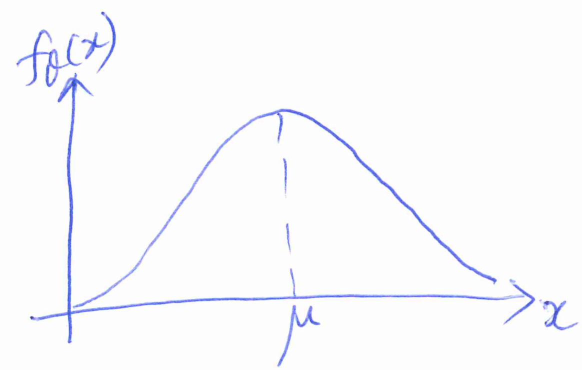
Some key model distributions

Discrete distributions

- Bernoulli(θ), $X \sim \text{Bernoulli}(\theta)$, $E[X] = P[X=1] = \theta$
- Binomial (n, θ), $E[X] = n\theta$, $\text{var}[X] = n\theta(1-\theta)$
- Poisson (θ), $E[X] = \text{var}[X] = \theta$

Continuous distributions

- Exponential (θ)
- Uniform $[a, b]$
- Normal (μ, σ^2), $\mu = E(X)$, $\sigma^2 = \text{var}(X)$,



Law of large numbers and central limit theorem

Suppose the rvs $X_1, \ldots X_n$ are **independently and identically** distributed (i.i.d.) as X where $\mu = E(X)$ and $\sigma^2 = \mathrm{var}(X)$.

Population : $X \sim f_\theta(x)$

$\mu = E[X]$, $\sigma^2 = var[X]$

$\downarrow$

$X_1, X_2, \ldots, X_n$

- $X_1, \ldots, X_n$ represent a **random sample** of size n from the population represented by X.

- Define **sample sum**, $T = \sum_{i=1}^{n} X_i$
- (verify) $E[T] = n\mu$, $\text{var}[T] = n\sigma^2$
- Define **sample average**, $\overline{X} = T/n$
- (verify) $E[\bar{X}] = \mu$, $\text{var}[\bar{X}] = \frac{\sigma^2}{n}$

always true.

distribution of sample-based quantities is also called "sampling distribution".

Law of large numbers (LLN)