Data Analytics
EEE 4774 & 6777

Module 2

Frequentist vs. Bayesian Probability

Spring 2022

Uncertainty & Probability

- ***Uncertainty*** in data:
  - inherent in the observed physical process (e.g., voltage measurement in power grid, # customers in a market)
  - noise in measurement (e.g., hardware/software limitations)
  - finite data size (i.e., lack of access to the entire population)

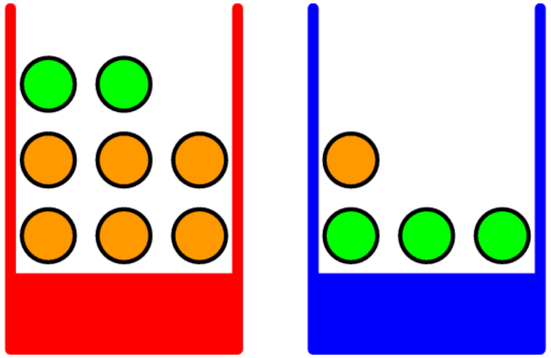
- ***Probability***:
  - a consistent framework for quantification and manipulation of uncertainty
  - helps in decision making (e.g., our brains)

Probability & Statistics in Data Science

- Compute statistics such as mean and variance to get insight
  - E.g., mean and standard deviation of age or height data in this class
- Build probabilistic models of data to use statistics in a systematic way
  - Classify data instances
  - Predict future values
  - Estimate missing values
  - Generate realistic (simulated) data
  - Assess the confidence in decisions
    - class probabilities, confidence intervals for predictions

Frequentist probability

- Frequency of observations
  - *marginal probability* $p(Y = o) = n_o/n$
  - *joint probability* $p(X = r, Y = g) = n_{rg}/n$
  - *conditional probability* $p(Y = g|X = b) = n_{bg}/n_b$



		X: r	X: b	
Y	g	n_{rg}	n_{bg}	n_g
	o	n_{ro}	n_{bo}	n_o
		n_r	n_b	

- Sum rule: $p(X) = \sum_Y p(X, Y)$

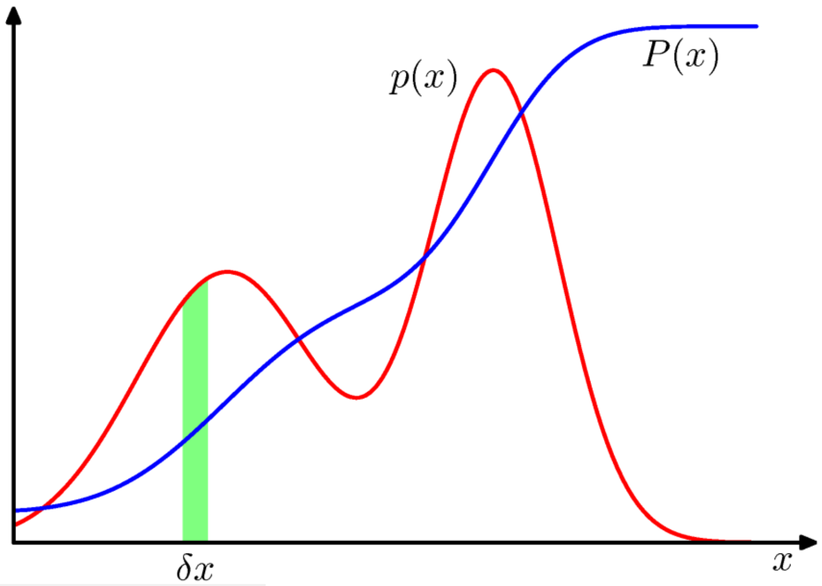
- Product rule: $p(X, Y) = p(Y|X)p(X)$

- ***Bayes Theorem:*** $p(X|Y) = \frac{p(X, Y)}{p(Y)} = \frac{p(Y|X)p(X)}{\sum_X p(Y|X)p(X)}$

$p(X = b|Y = g) =?$

Probability

- $p(x)$: Probability density/mass function of a continuous/discrete variable



$p(x \,\epsilon\, (a,b)) = \int_a^b p(x)\, dx$	$p(x) \ge 0$	$\int_{-\infty}^{\infty} p(x)\, dx = 1$
$P(y) = \int_{-\infty}^{y} p(x) dx$	$p(\mathbf{x}) \ge 0$	$\iiint p(\mathbf{x})\, d\mathbf{x} = 1$
$p(x) = \int p(x,y)\, dy$	$p(x,y) = p(y \mid x) p(x)$	$E[f(x)] = \sum_x f(x) p(x)$
$E[f(x)] = \int f(x) p(x) dx$	$E_x[f(x,y)] = \int f(x,y) p(x,y) dx$	$E_{x \mid y}[f(x,y) \mid y] = \int f(x,y) p(x \mid y) dx$
$Var[f]$ $= E[(f(x) - E[f(x)])^2]$ $= E[f(x)^2] - E[f(x)]^2$	$Cov[x,y]$ $= E_{x,y}[(x - E[x])(y - E[y])]$ $= E_{x,y}[xy] - E[x]E[y]$	$Cov[\mathbf{x},\mathbf{y}]$ $= E_{\mathbf{x},\mathbf{y}}\left[(\mathbf{x} - E[\mathbf{x}])(\mathbf{y} - E[\mathbf{y}])^T\right]$ $= E_{\mathbf{x},\mathbf{y}}\left[\mathbf{x}\mathbf{y}^T\right] - E[\mathbf{x}]E[\mathbf{y}^T]$

Bayesian Probability

- ***Classical/Frequentist*** interpretation of probability ~ frequencies of ***repeatable*** events

- ***Bayesian probability*** ~ a quantification of uncertainty
  - repeatable and ***non-repeatable*** events, e.g., the probability of *a dragon flying through the window*
  - ***update with evidence***, e.g., it is shown that there exist dragons in Florida, and there are small ones that can fit through a window.

$$p(\mathbf{x}|D) = \frac{p(D|\mathbf{x})\, p(\mathbf{x})}{p(D)}$$

posterior α *likelihood* x *prior*

Bayesian Probability

$$p(\mathbf{x}|D) = \frac{p(D|\mathbf{x})\, p(\mathbf{x})}{p(D)}$$

posterior α *likelihood* x *prior*

- ***Prior probability*** is not an arbitrary choice, *reflects common sense* (or uninformative)

- ***Challenge:*** for predictions and model comparison, marginalization typically difficult!

$$p(D) = \iiint p(D|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}$$

Bayesian vs. Frequentist

	Bayesian	Frequentist
Likelihood	fixed data, random parameters	random data, fixed parameters
Model Selection	training data: evidence (Occam's razor)	training + validation data: cross validation (may be computationally cumbersome)
Regularization	naturally provided by ***prior*** (prevents overfitting)	needs additional penalty
Accuracy	naturally provided by ***posterior*** (quality evaluation)	needs additional techniques (confidence interval, bootstrap)

- Bayesian ***prior*** *may not be realistic*, but *more and more training data* decreases the effect of prior

- Advances in *computational power*, as well as *techniques for computing posterior & marginal*

(e.g., *sampling techniques* such as MCMC, and *approximate inference* such as variational Bayes)

promote Bayesian approach, enable its use in *Big Datasets*.