Class Activity [7 minute]:
Most Basic Bayesian Interview Question

> What's the chance you have the disease if you tested postive?

- What do you need to know to answer this?

$$Pr(D|+) = ?$$

Class Activity [10 minutes]:
Most Basic Bayesian Interview Question

> What's the chance you have the disease if you tested postive?

- What do you need to know to answer this?

$$\begin{array}{rcr} Pr(D) = p & \implies & Pr(\text{not } D) = 1 - p \\ Pr(+|D) = q & \implies & Pr(-|D) = 1 - q \\ Pr(+|\text{not } D) = q' & \implies & Pr(-|\text{not } D) = 1 - q' \end{array}$$

$$\color{red}{Pr(D|+) = ?}$$

Answer Review [4 minutes]
Most Basic Bayesian Interview Question

Law of Total Probability

$$Pr(+) = \sum_{x \in \mathcal{X}} Pr(+, x) =$$

Answer Review [4 minutes]
Most Basic Bayesian Interview Question

Law of Total Probability

$$
\begin{aligned}
Pr(+) &= \sum_{x \in \mathcal{X}} Pr(+, x) = Pr(+, D) + Pr(+, \text{not } D)\\
&= Pr(+|D)Pr(D) + Pr(+|\text{not } D)Pr(\text{not } D) = qp + q'(1-p)
\end{aligned}
$$

Answer Review [4 minutes]
Most Basic Bayesian Interview Question

Law of Total Probability

$$
\begin{aligned}
Pr(+) &= \sum_{x \in \mathcal{X}} Pr(+, x) = Pr(+, D) + Pr(+, \text{not } D)\\
&= Pr(+|D)Pr(D) + Pr(+|\text{not } D)Pr(\text{not } D) = qp + q'(1-p)
\end{aligned}
$$

Bayes Theorem

$$Pr(D|+) =$$

Answer Review [4 minutes]
Most Basic Bayesian Interview Question

Law of Total Probability

$$
\begin{aligned}
Pr(+) &= \sum_{x\in\mathcal{X}} Pr(+,x) = Pr(+,D) + Pr(+,\text{not } D)\\
&= Pr(+|D)Pr(D) + Pr(+|\text{not } D)Pr(\text{not } D) = qp + q'(1-p)
\end{aligned}
$$

Bayes Theorem

$$
Pr(D|+) = \frac{Pr(+|D)Pr(D)}{Pr(+)} = \frac{Pr(+|D)Pr(D)}{Pr(+|D)Pr(D) + Pr(+|\text{not } D)Pr(\text{not } D)} = \frac{qp}{qp + q'(1}
$$

Bayesian Versus Frequentist Philosophy [2 minutes]

- ***FREQUENTIST***:
  -
- ***BAYESIAN***:
  -
  -

Bayesian Versus Frequentist Philosophy [5 minutes]

- ***FREQUENTIST***:
  Parameter θ in a **likelihood** $f_\theta(x)$ is a fixed value to be estimated
  - No interesting probability statements to be made about **parameters**
- ***BAYESIAN***:
  The **uncertainty** in θ can be modeled as (**posterior**) **distribution** $p(\theta|x)$
  - All statements about **parameters** are probability statements

- **Bayes theorem** coherently assess **parameter uncertainty** as a **posterior** $p(\theta|x)$ given **likelihood** $f_\theta(x) = p(x|\theta)$ and **prior** $p(\theta)$

Bayesian Terminology [10 minutes]

$$
\begin{array}{rl}
\text{Likelihood} & p(x|\theta) \quad \text{often written as } f_\theta(x) \\
\text{Prior} & p(\theta) \\
\text{Posterior} & p(\theta|x) = \overset{\text{Bayes Theorem}}{\frac{p(x|\theta)p(\theta)}{p(x)}} = \frac{p(x,\theta)}{p(x)} \\
\text{Marginal Likelihood} & p(x) = \underbrace{\int p(x|\theta)p(\theta)d\theta}_{\text{Law of Total Probability}} \\
\text{Join Distribution} & p(\theta, x) \\
\text{The Posterial is always} & p(\theta|x) \propto p(\theta, x) \\
& \qquad\qquad \text{proportional to the joint}
\end{array}
$$

Bayesian/Sequential Learning [9 minutes]

$$
\text{Conditional Independence} \qquad p(x_2|\theta) = p(x_2|x_1, \theta) \;\text{ or }\; (x_2 \perp\!\!\!\perp x_1|\theta)
$$

$$
\begin{aligned}
p(\theta|x_2, x_1) &\propto p(x_2, x_1, \theta) \\
&= \underbrace{p(x_2|\theta, x_1)}\, p(x_1|\theta)p(\theta) \longleftarrow \text{chain rule} \\
&= \overbrace{\;p(x_2|\theta)\;}^{(x_2 \perp\!\!\!\perp x_1|\theta)}\; \underbrace{p(x_1|\theta)p(\theta)}_{\propto\, p(\theta|x_1)} \\
&\propto p(x_2|\theta)\underbrace{p(\theta|x_1)}_{\text{new prior}} \longleftarrow \text{previous posterior}
\end{aligned}
$$

A prior is updated to a posterior which becomes the new prior for the next data

Bayesian/Sequential Learning [2 minutes]

A prior is updated to a posterior which becomes the new prior for the next data

1. $p(\theta|x_1) \propto f(x_1|\theta)p(\theta)$
2. $p(\theta|x_1, x_2) \propto f(x_2|\theta)p(\theta|x_1)$
3. $p(\theta|x_1, x_2, x_3) \propto f(x_3|\theta)p(\theta|x_1, x_2)$

$\vdots$

Bayesian/Sequential Learning

A prior is updated to a posterior which becomes the new prior for the next data

1. $p(\theta|x_1) \propto f(x_1|\theta)p(\theta)$
2. $p(\theta|x_1, x_2) \propto f(x_2|\theta)p(\theta|x_1)$
3. $p(\theta|x_1, x_2, x_3) \propto f(x_3|\theta)p(\theta|x_1, x_2)$

$\vdots$

Is this different from Batch Updates? [5 minutes]

$$p(\theta|x_1, x_2, x_3) \propto f(x_1, x_2, x_3|\theta)p(\theta) \overset{(x_i \perp\!\!\!\perp x_j|\theta)}{=} \underbrace{f(x_1|\theta)f(x_2|\theta)f(x_3|\theta)p(\theta)}_{\left[\prod_{i=1}^{3} f(x_i|\theta)\right]p(\theta)}$$

Criticisms of Bayesian Analysis [3 minutes]

Criticisms of Bayesian Analysis [6 minutes]

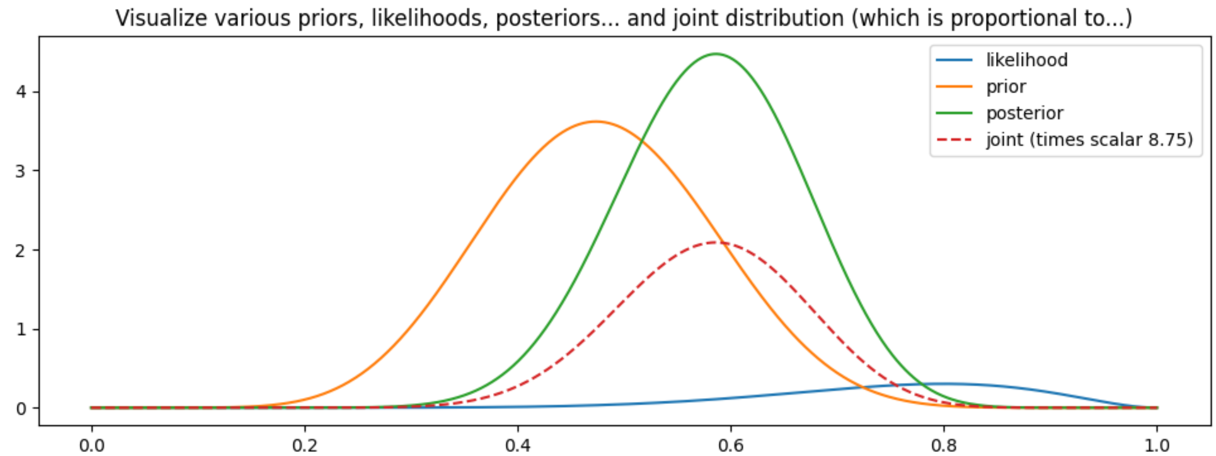
- "Your prior is subjective"
    - Response: "Your likelihood is subjective",
      "Your model selection is subjective", etc. Where does this stop...?
        - Affects **Frequentists** as well...
          "All models are wrong, some are useful."
- "Bayesian analysis is computationally expensive"
    - Response: In the end, a method is or is not computationally tractable. Faster methods may always be preferred...
        - There will be things you **can't** do unless you're ***Bayesian***

A first (toy) model: beta-binomial [15 minutes]

$$
\begin{align*}
p(x|\theta,n) &= \binom{n}{x}\theta^x(1-\theta)^{n-x} && [n \text{ given}]\\
p(\theta|\alpha,\beta) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} && \alpha \text{ and } \beta \text{ are called } \textbf{hyperparameters}\\
p(\theta|x) &= \;?\\
\text{Hint: } p(\theta|x) &\propto p(\theta,x)\\
&= p(x|\theta)p(\theta)\\
&= \binom{n}{k}\theta^x(1-\theta)^{n-x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\\
&\propto \theta^x(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}\\
&= \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \Rightarrow \text{ What distribution?}
\end{align*}
$$

The "Classic Bayes" Figure [10 minutes]

In [ ]:
```
import numpy as np; from scipy import stats; import matplotlib.pyplot as plt
n,x=10,8; alpha,beta=10,11; parameter_domain = np.linspace(0,1,1001)
likelihood = stats.binom.pmf(n=n, k=x, p=parameter_domain)
prior = stats.beta(a=alpha,b=beta).pdf(parameter_domain)
c = 8.75; joint = c*likelihood*prior # elementwise multiplication
posterior = stats.beta(a=x+alpha,b=n-x+beta).pdf(parameter_domain)
plt.figure(figsize=(12,4)); plt.plot(parameter_domain, likelihood, label="likelihoo
```



Homework #1: AB-testing and the Multi-Armed Bayesian Bandit

You have three choices... choose wisely, my friends...

Option	The "Red" one	The "Blue" one	The "Other" one
Unknown Probability of Success	θ_A	θ_B	θ_C

$$p(\theta_j|x_j, n_j) \propto \theta^{x_j+\alpha_j-1}(1-\theta_j)^{n-x+\beta_j-1} \Rightarrow \text{ What distribution?}$$

- Try one out, and collect that data update...
  - What's the data?
  - What's the update for the posterior in question?
- Which one of the three choices will you try out? How will you choose?
- Hints: You can use *simulation* to find out the *relative belief* (i.e., probability) that each of the choices is the best. Posterior distributions characterize your beliefs about the parameters θ_A, θ_B and θ_C. What can you learn by repeatedly sampling values from the posterior distribution while comparing the values of each triplet? If you know the chances that A, B, and C are the best choice, how could you balance ***exploration versus exploitation*** when choosing which of the possible options to collect the next data point on next?