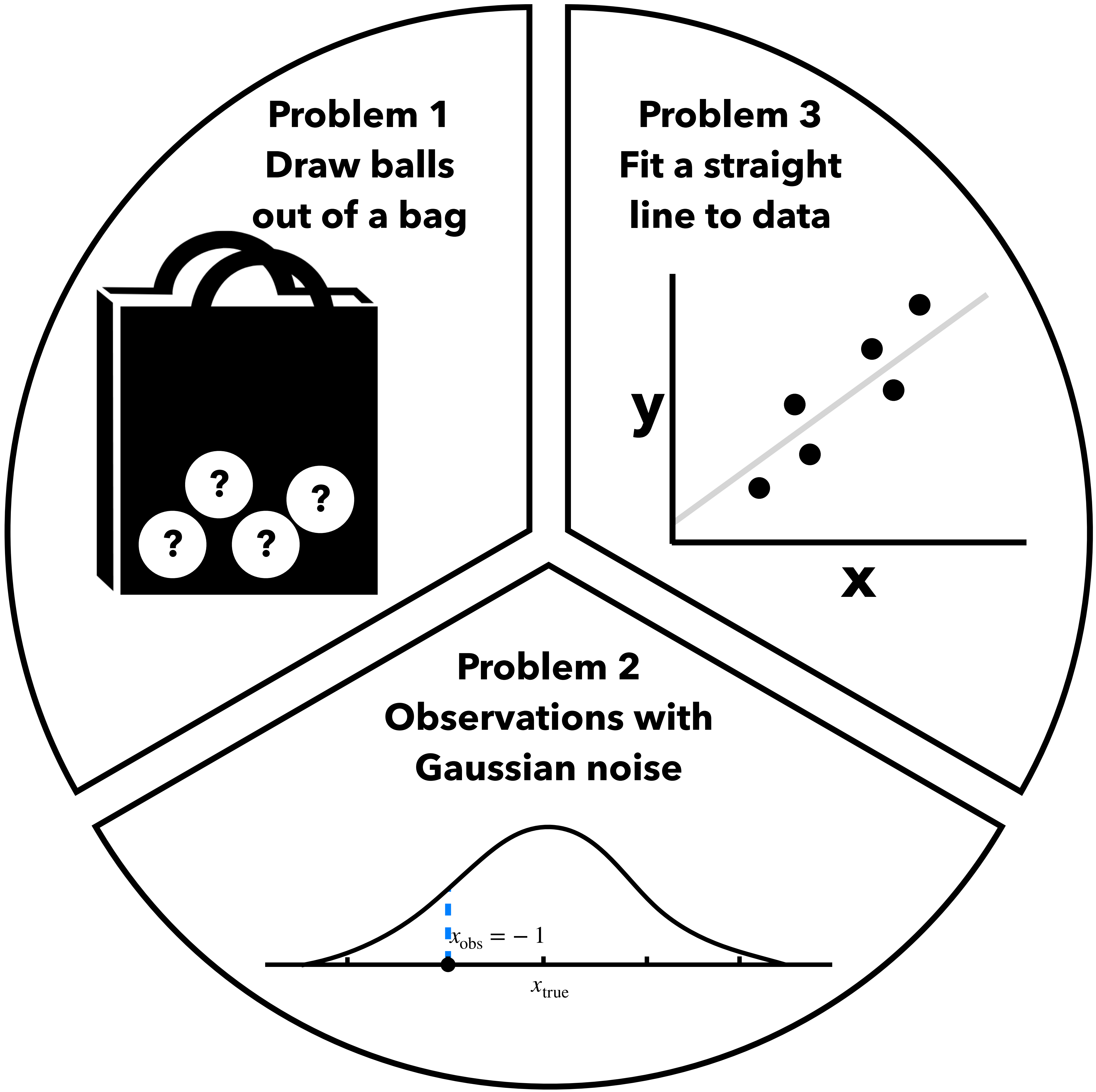
Introduction to Bayesian Statistics

LSSTC Fellowship Program Session 16

Jiayin Dong, Flatiron Research Fellow
Center for Computational Astrophysics, Flatiron Institute
9/19/2022

$$p(\theta \mid D, H) = \frac{p(D \mid \theta, H)\,p(\theta \mid H)}{p(D \mid H)}$$

Bayes' theorem

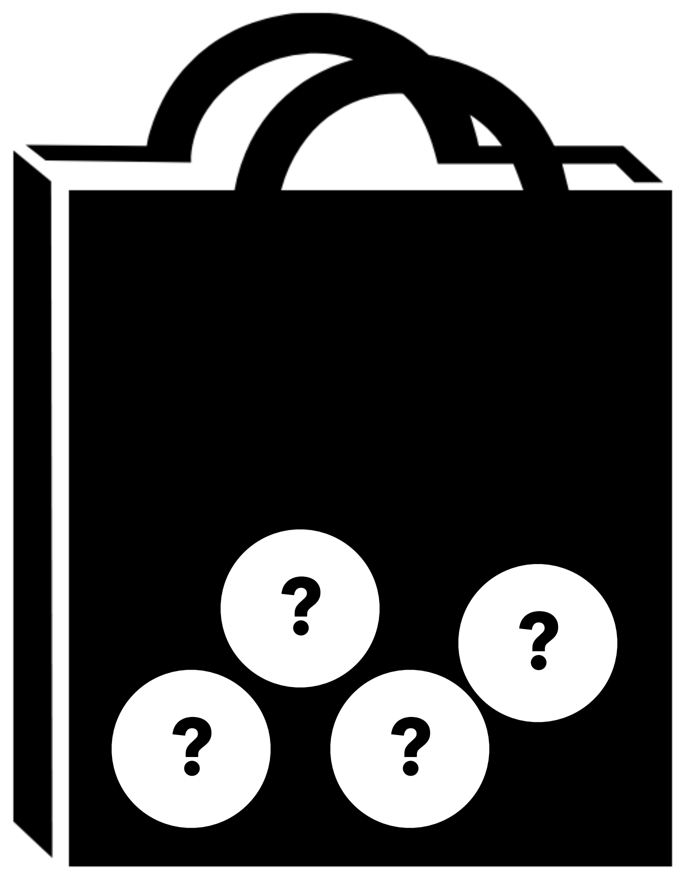


Problem 1
Draw balls out of a bag
?
?
?
?
Problem 3
Fit a straight line to data
y
x
Problem 2
Observations with Gaussian noise
$x_{obs} = -1$
x_{true}

We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

Q: What are all possible combinations of balls?

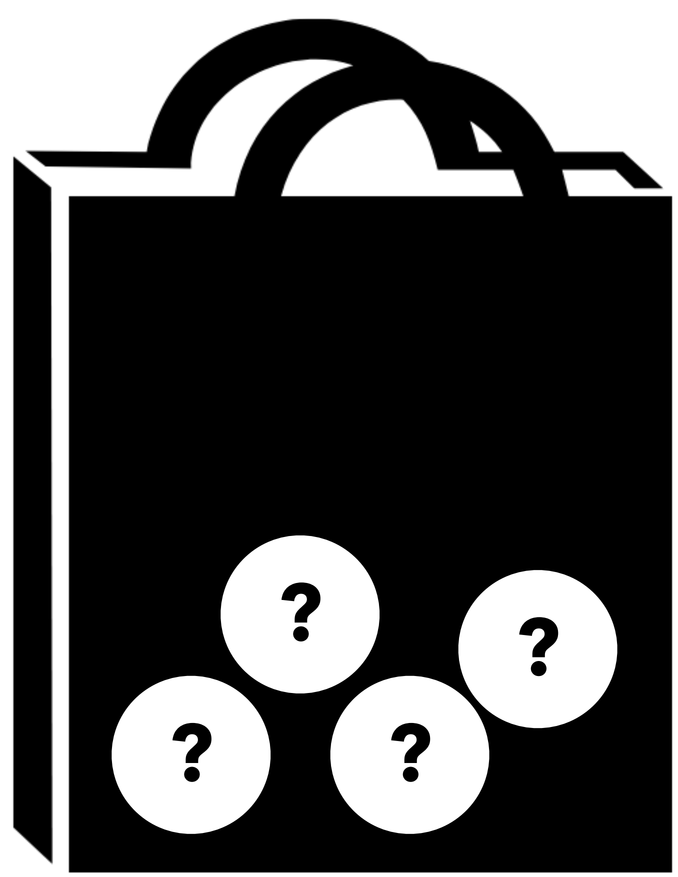


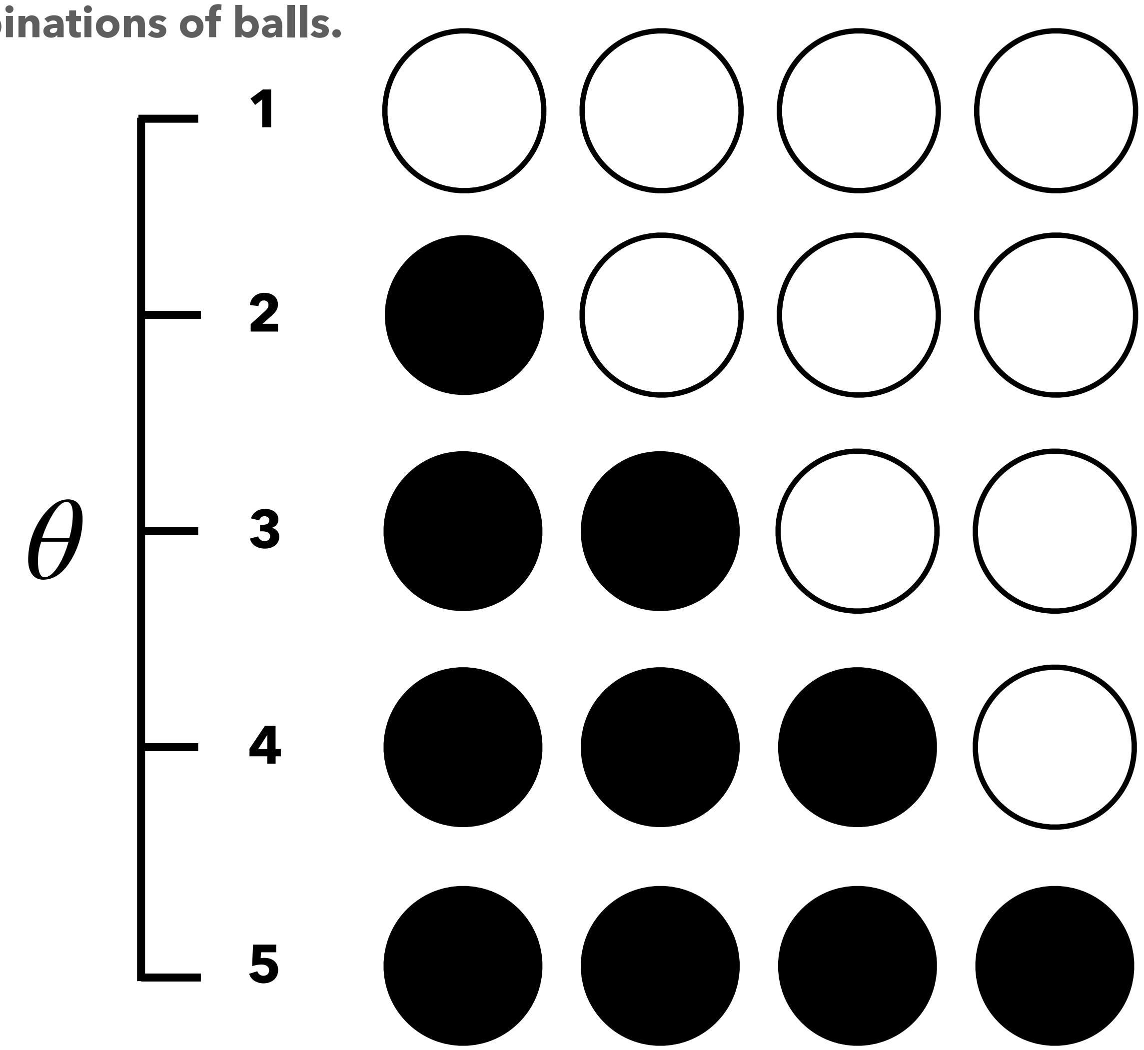
We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

Q: What are all possible combinations of balls?

A Parameter of Interest θ describes all combinations of balls.

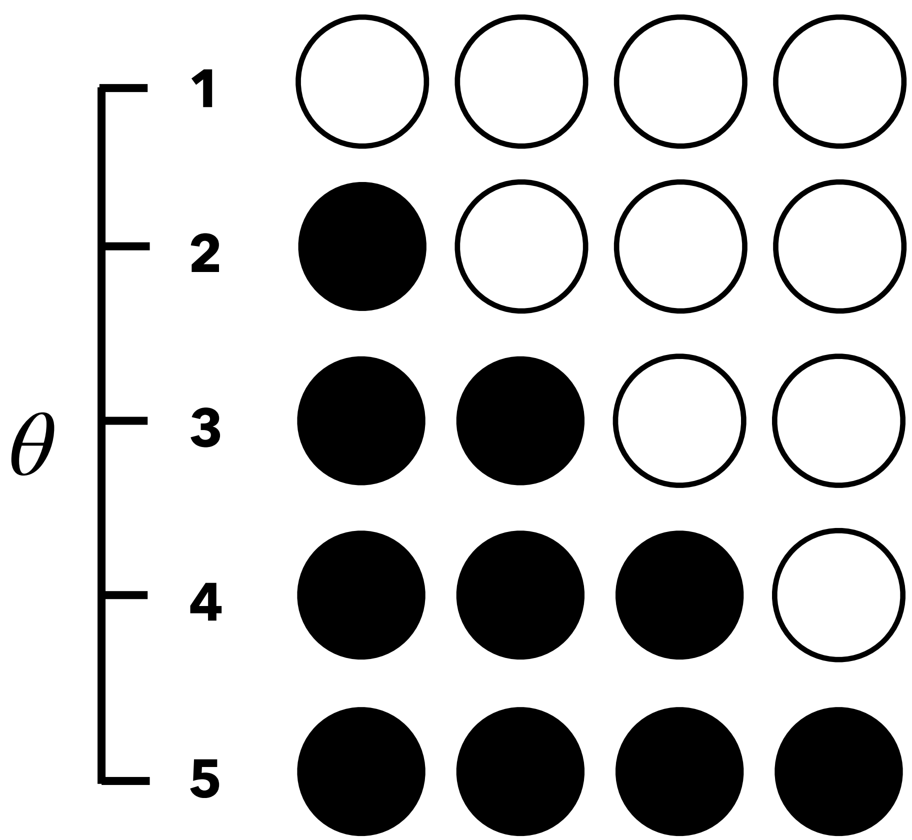




We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

Q: What are all possible combinations of balls?



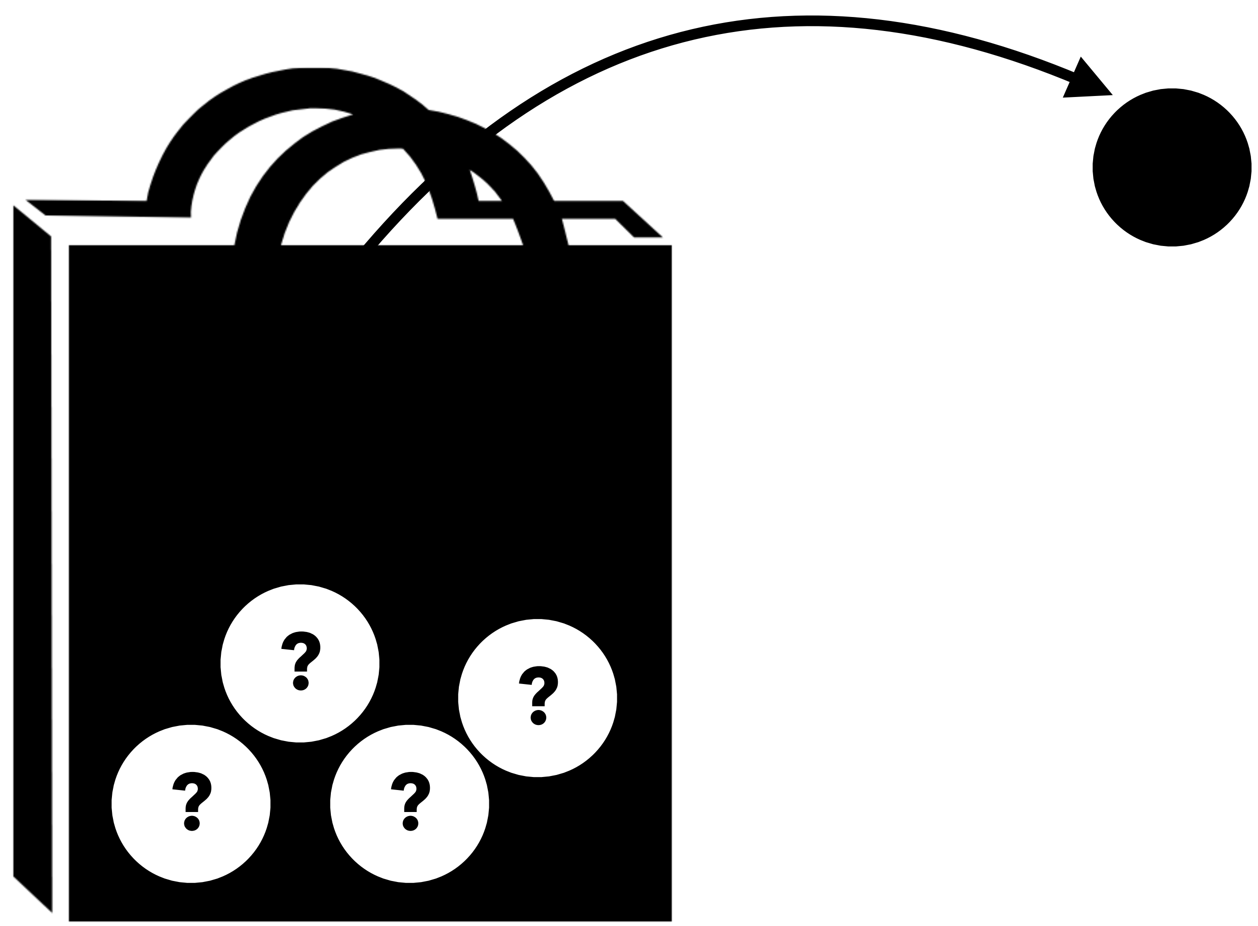
Uninformative guess

Note: $p(\theta)$ **is a discrete distribution.**

We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

We draw 1 ball from the bag and it's black.



We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

We draw 1 ball from the bag and it's black.

Q: What is the probability of the observation at each combination?

We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

We draw 1 ball from the bag and it's black.

Q: What is the probability of the observation at each combination?

θ		Probability of drawing a black ball	
1	○○○○	0/4	Pr("●"|○○○○) = 0
2	●○○○	1/4	Pr("●"|●○○○) = 0.25
3	●●○○	2/4	Pr("●"|●●○○) = 0.5
4	●●●○	3/4	Pr("●"|●●●○) = 0.75
5	●●●●	4/4	Pr("●"|●●●●) = 1

We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

We draw 1 ball from the bag and it's black.

Q: What is the probability of the observation at each combination?

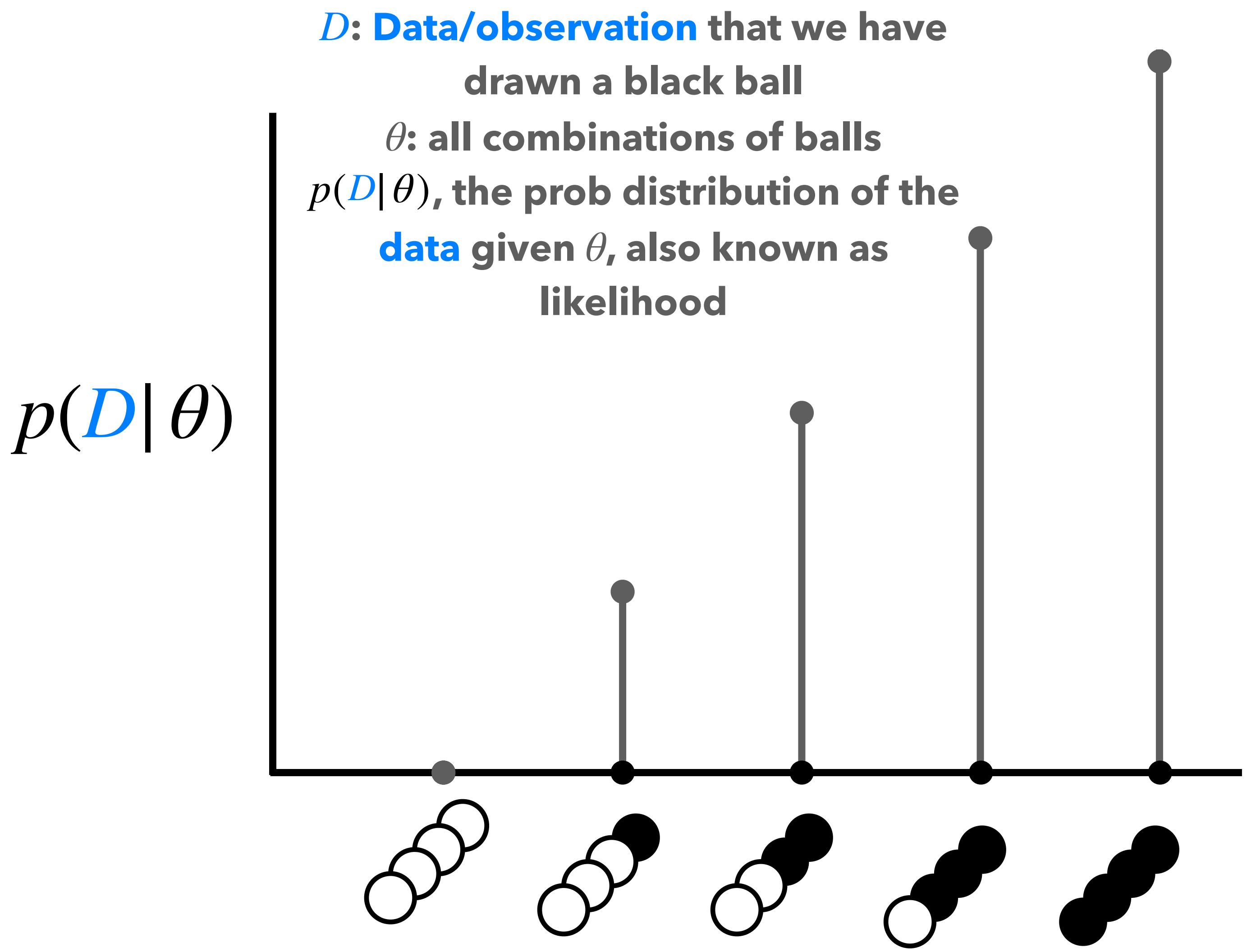
Pr("●"|○○○○) = 0

Pr("●"|●○○○) = 0.25

Pr("●"|●●○○) = 0.5

Pr("●"|●●●○) = 0.75

Pr("●"|●●●●) = 1

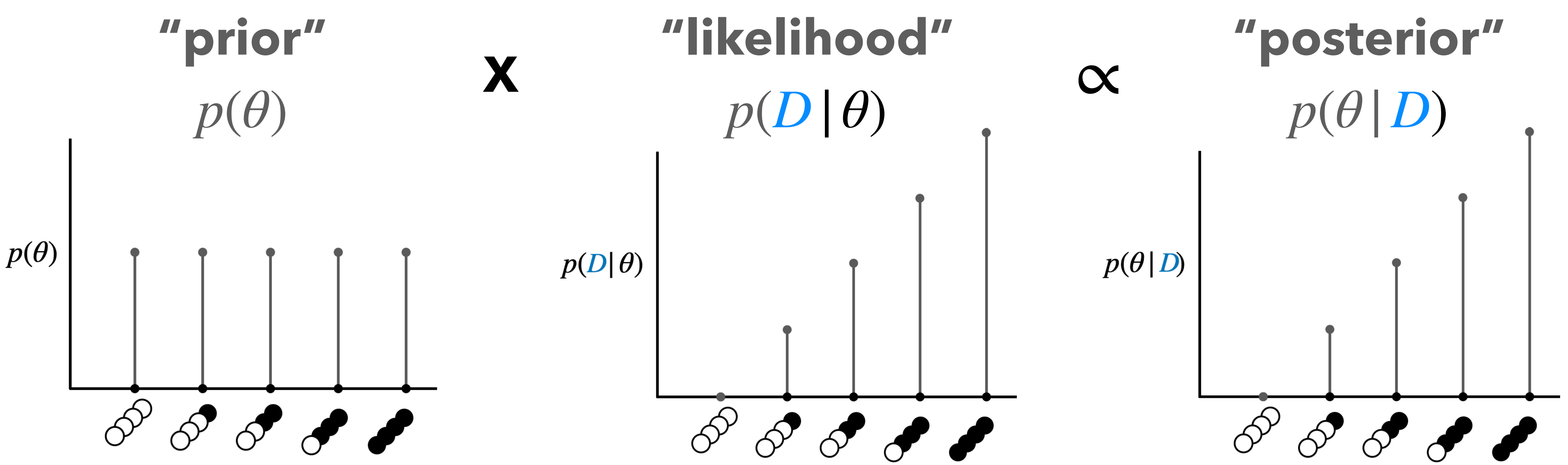


We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

We draw 1 ball from the bag and it's black.

Q: What is the updated probability distribution of θ?



We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

Draws with replacement. Draw 1: black; Draw 2: white.

Q: What is the posterior distribution of θ given the data?

We have a bag containing 4 balls.

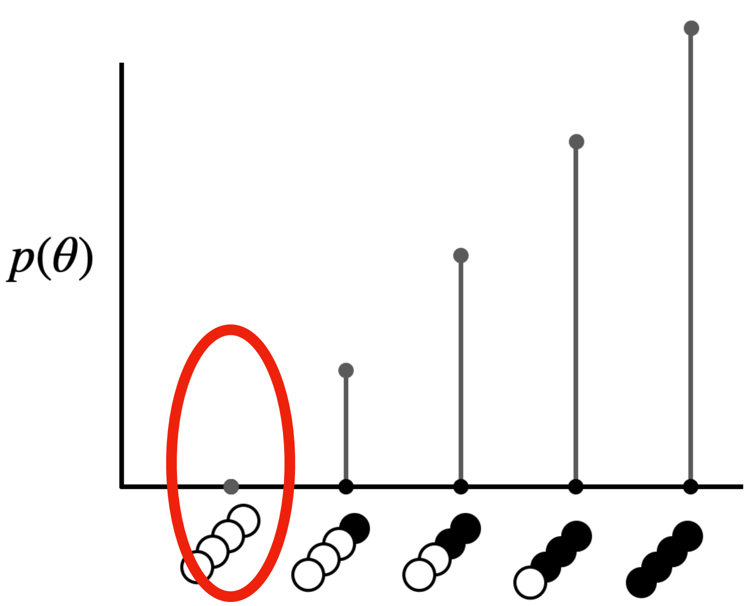
Each ball has two possible colors: black and white.

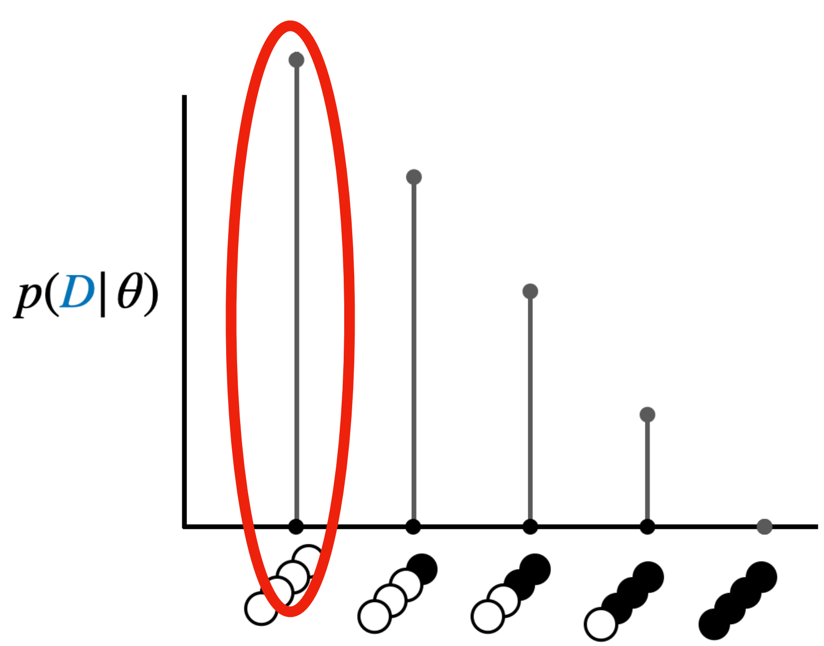
Draws with replacement. Draw 1: black; Draw 2: white.

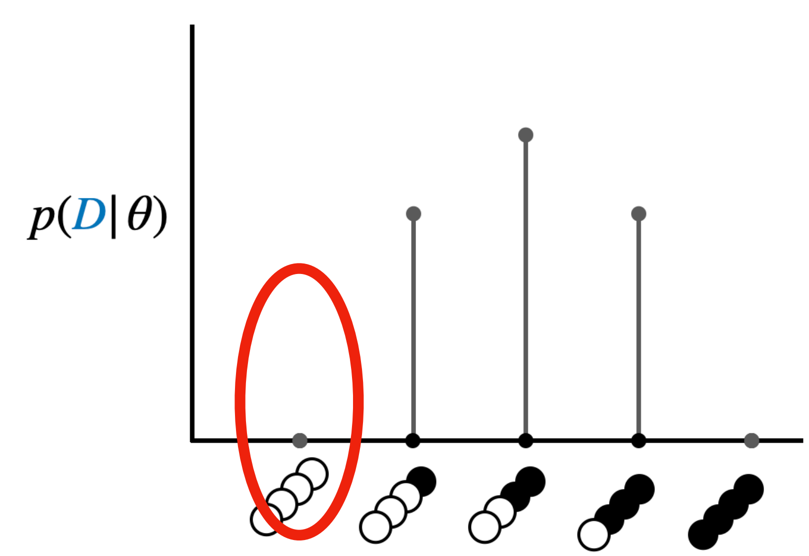
Q: What is the posterior distribution of θ given the data?

Approach 1: Use the posterior derived from Draw 1 as a prior.

"prior" $p(\theta)$ x "likelihood" $p(D \mid \theta)$ $\propto$ "posterior" $p(\theta \mid D)$







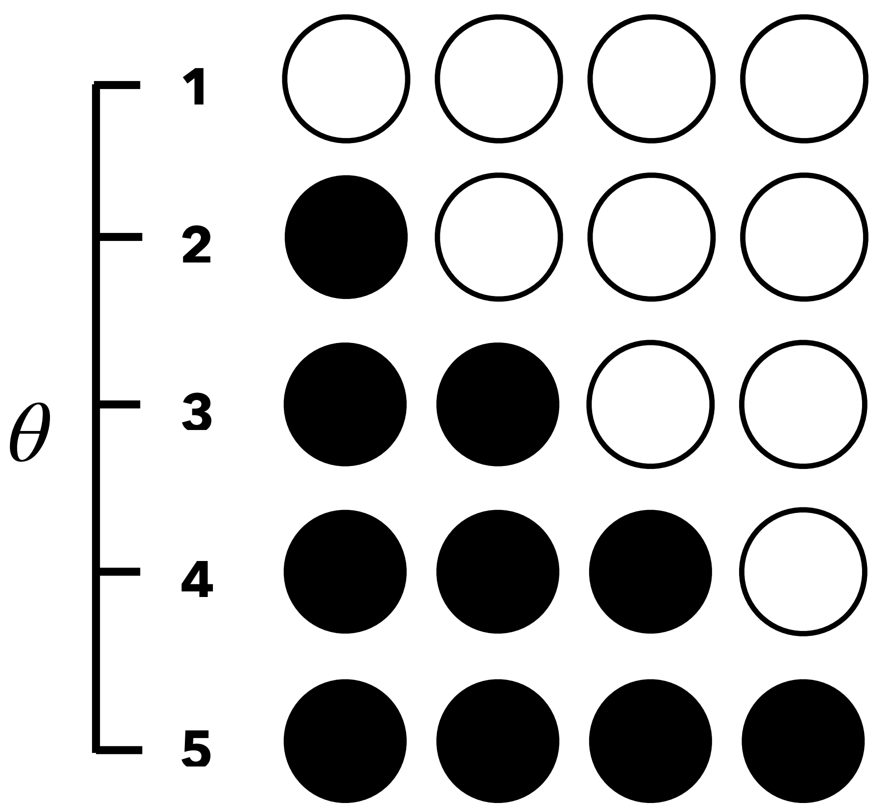
We have a bag containing 4 balls.

Each ball has two possible colors: black and white.

Draws with replacement. Draw 1: black; Draw 2: white.

Q: What is the posterior distribution of θ given the data?

Approach 2: Calculate the likelihood function of two observations.



Pr("●○"|○○○○) = 0

Pr("●○"|●○○○) = 3/16

Pr("●○"|●●○○) = 4/16

Pr("●○"|●●●○) = 3/16

Pr("●○"|●●●●) = 0

We have a bag containing 4 balls.

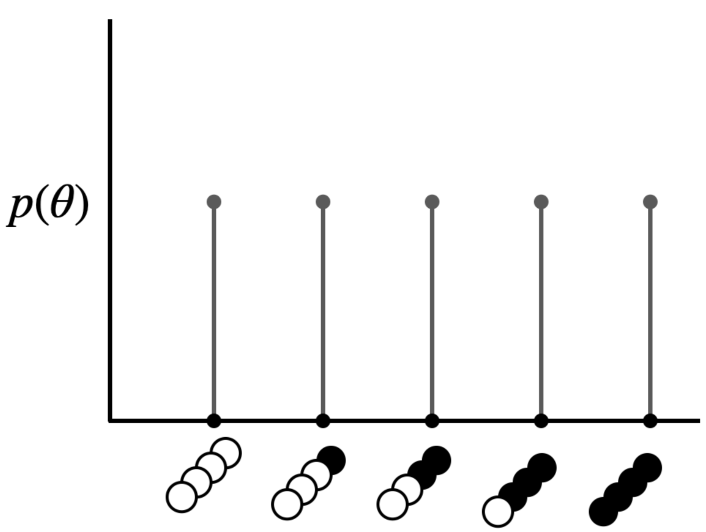
Each ball has two possible colors: black and white.

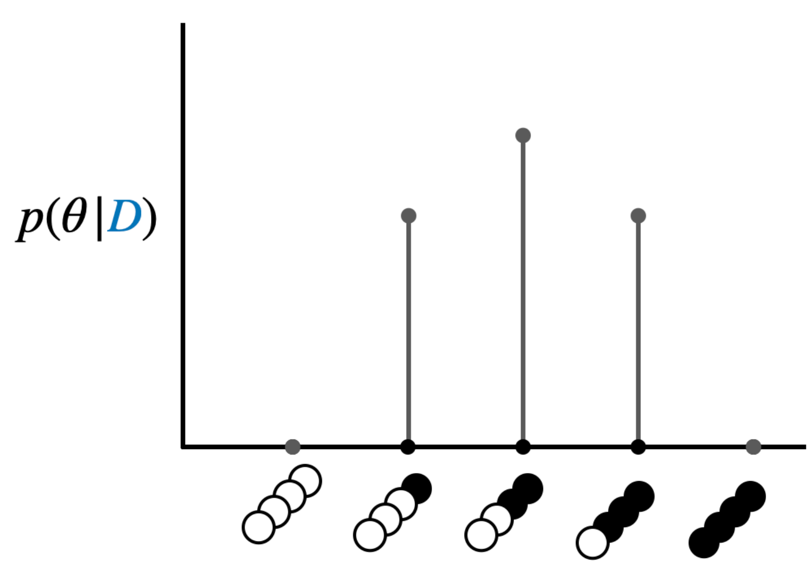
Draws with replacement. Draw 1: black; Draw 2: white.

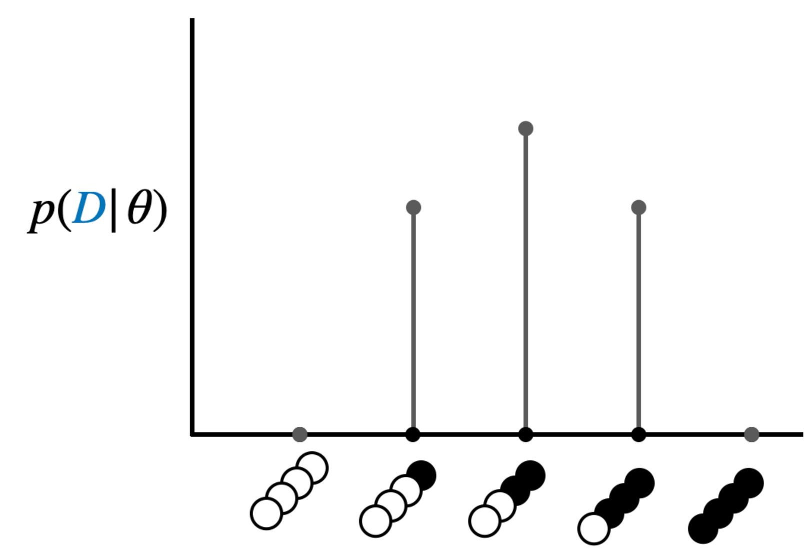
Q: What is the posterior distribution of θ given the data?

Approach 2: Calculate the likelihood function of two draws.

"prior"		"likelihood"		"posterior"
$p(\theta)$	X	$p(D \mid \theta)$	$\propto$	$p(\theta \mid D)$





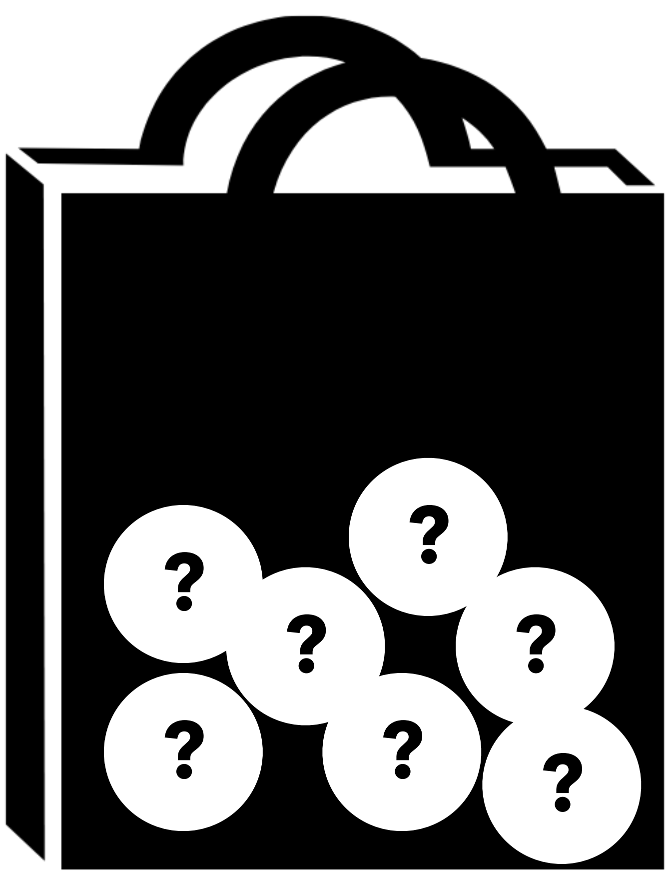


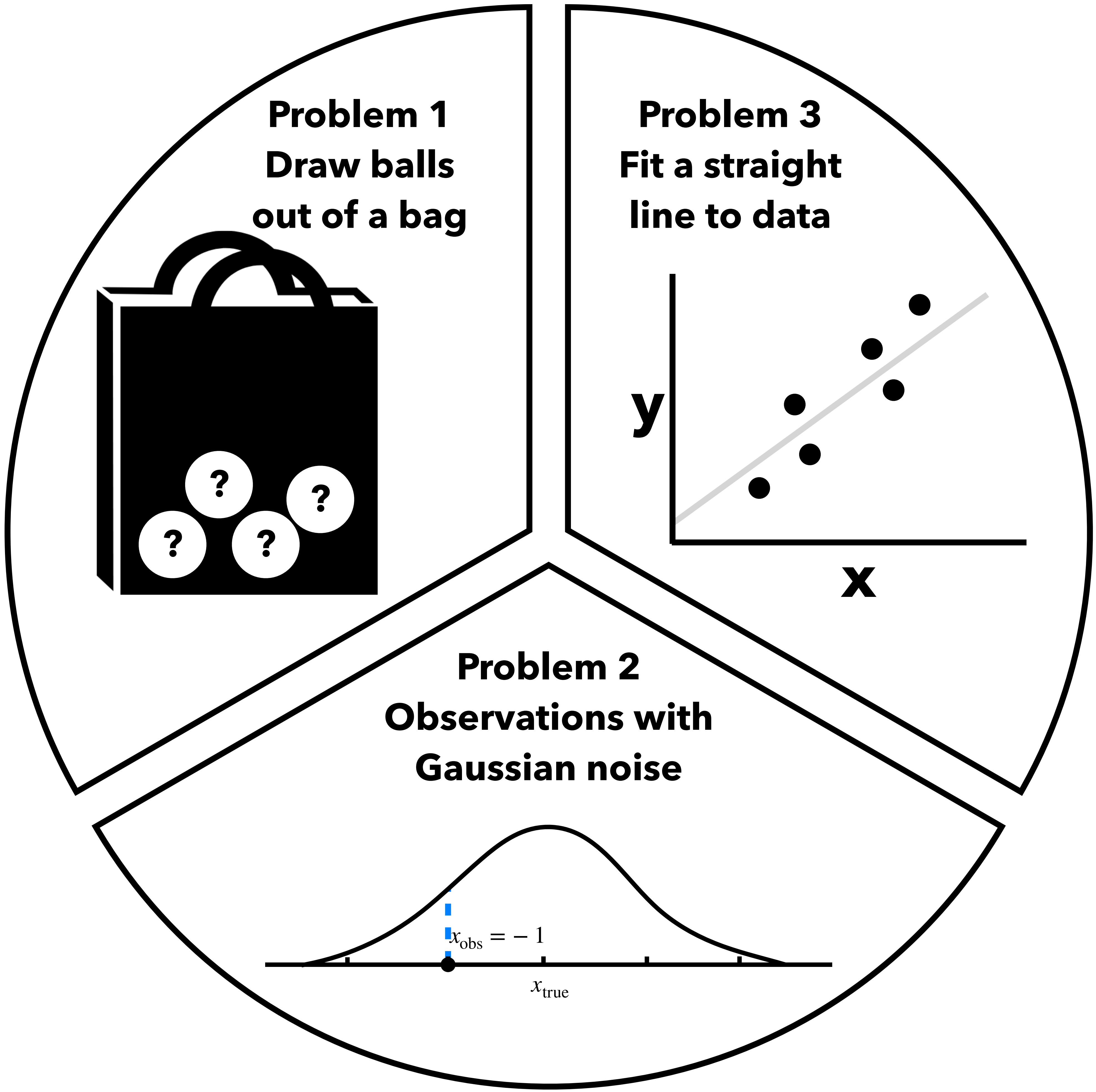
We have a bag containing *infinity* number of balls.

Each ball has two possible colors: black and white.

Made 10 draws with replacement. Observed 6 blacks, 4 whites

Q: What is the posterior distribution of θ given the data?




Problem 1
Draw balls out of a bag
?
?
?
?
Problem 3
Fit a straight line to data
y
x
Problem 2
Observations with Gaussian noise
$x_{obs} = -1$
x_{true}

We have a parameter of interest, x.

We made an observation of x, but the observation was noisy.

Instead of the "true" value of x, x_obs was observed.

Q: How do we infer x from x_obs?

$$x_{\text{obs}} = x_{\text{true}} + \mathcal{N}(0, \sigma^2)$$

$$p(\theta \mid D, H) = \frac{p(D \mid \theta, H)\,p(\theta \mid H)}{p(D \mid H)}$$

Bayes' theorem

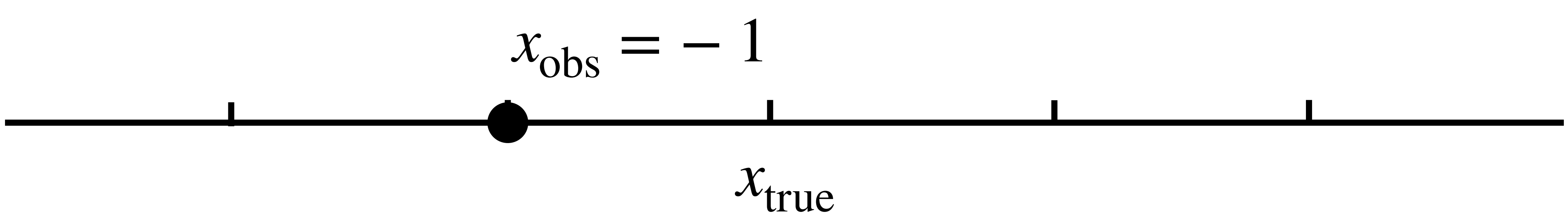
We have a parameter of interest, x.

We made an observation of x, but the observation was noisy.

Instead of the "true" value of x, x_obs was observed.

Q: How do we infer x from x_obs?

$$x_{\text{obs}} = x_{\text{true}} + \mathcal{N}(0, \sigma^2)$$



We have a parameter of interest, x.

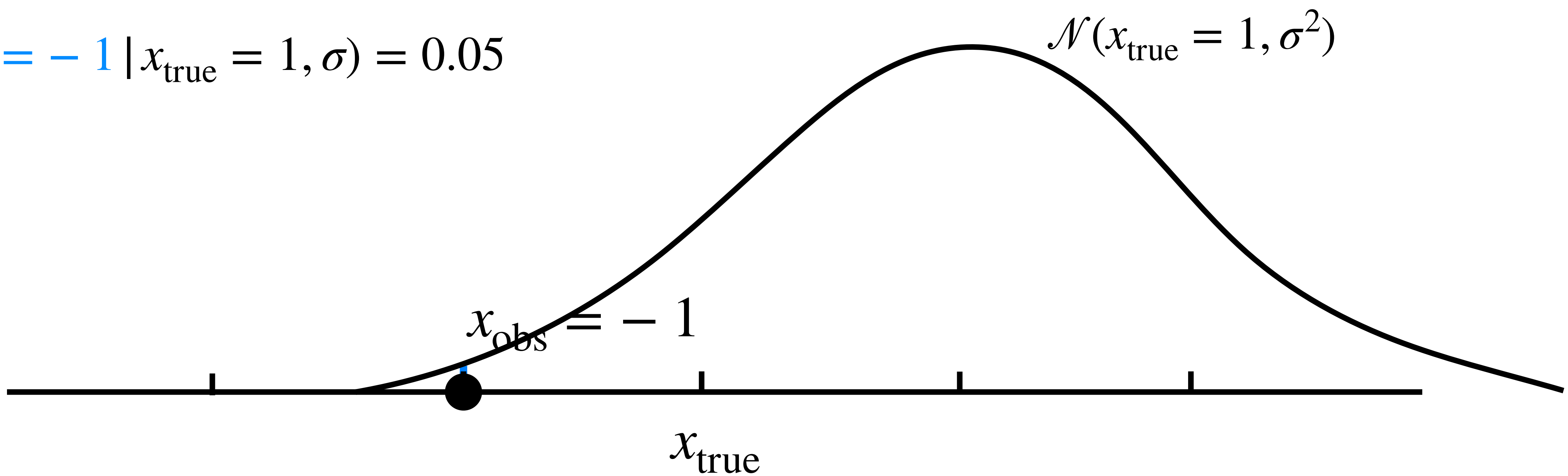
We made an observation of x, but the observation was noisy.

Instead of the "true" value of x, x_obs was observed.

Q: How do we infer x from x_obs?

$$x_{\text{obs}} = x_{\text{true}} + \mathcal{N}(0, \sigma^2)$$

$p(x_{\text{obs}} = -1 \,|\, x_{\text{true}} = 1, \sigma) = 0.05$



We have a parameter of interest, x.

We made an observation of x, but the observation was noisy.

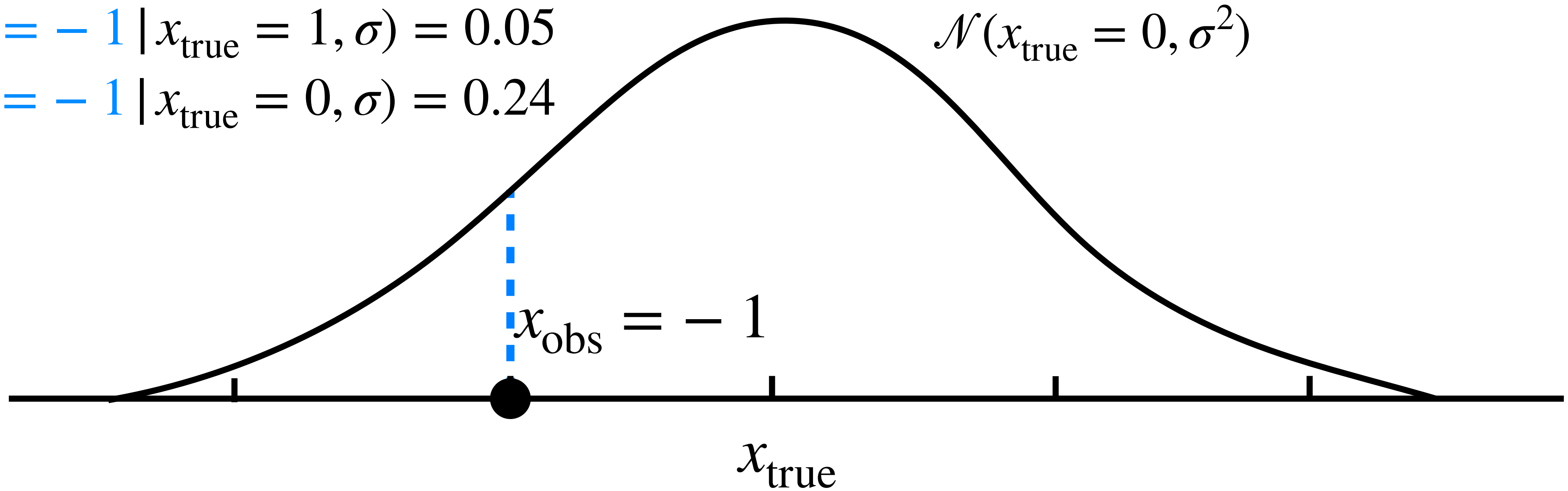
Instead of the "true" value of x, x_obs was observed.

Q: How do we infer x from x_obs?

$$x_{\text{obs}} = x_{\text{true}} + \mathcal{N}(0, \sigma^2)$$

$p(x_{\text{obs}} = -1 \,|\, x_{\text{true}} = 1, \sigma) = 0.05$

$p(x_{\text{obs}} = -1 \,|\, x_{\text{true}} = 0, \sigma) = 0.24$



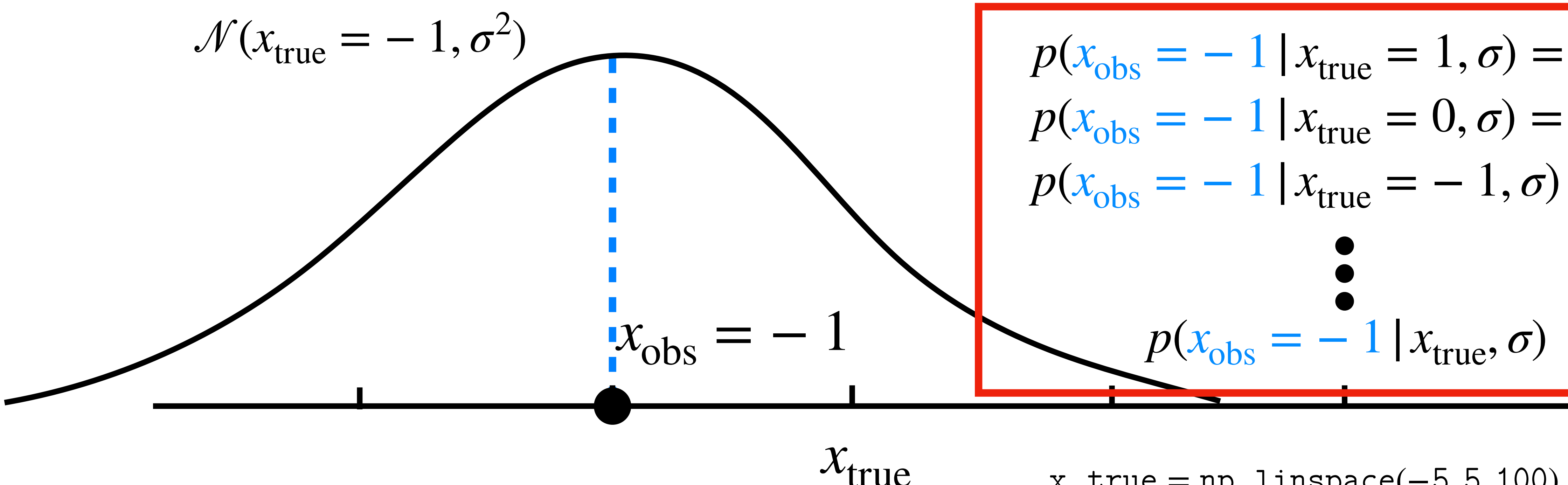
We have a parameter of interest, x.

We made an observation of x, but the observation was noisy.

Instead of the "true" value of x, x_obs was observed.

Q: How do we infer x from x_obs?

$$x_{\text{obs}} = x_{\text{true}} + \mathcal{N}(0, \sigma^2)$$



This is the likelihood function.

$$
\boxed{
\begin{aligned}
&p(x_{\text{obs}} = -1 \,|\, x_{\text{true}} = 1, \sigma) = 0.05 \\
&p(x_{\text{obs}} = -1 \,|\, x_{\text{true}} = 0, \sigma) = 0.24 \\
&p(x_{\text{obs}} = -1 \,|\, x_{\text{true}} = -1, \sigma) = 0.40 \\
&\vdots \\
&p(x_{\text{obs}} = -1 \,|\, x_{\text{true}}, \sigma)
\end{aligned}
}
$$

$\texttt{x\_true} = \texttt{np.linspace}(-5, 5, 100)$

$\texttt{norm.pdf}(\texttt{x\_obs}, \texttt{loc} = \texttt{x\_true}, \texttt{scale} = \sigma)$

We have a parameter of interest, x.

We made an observation of x, but the observation was noisy.

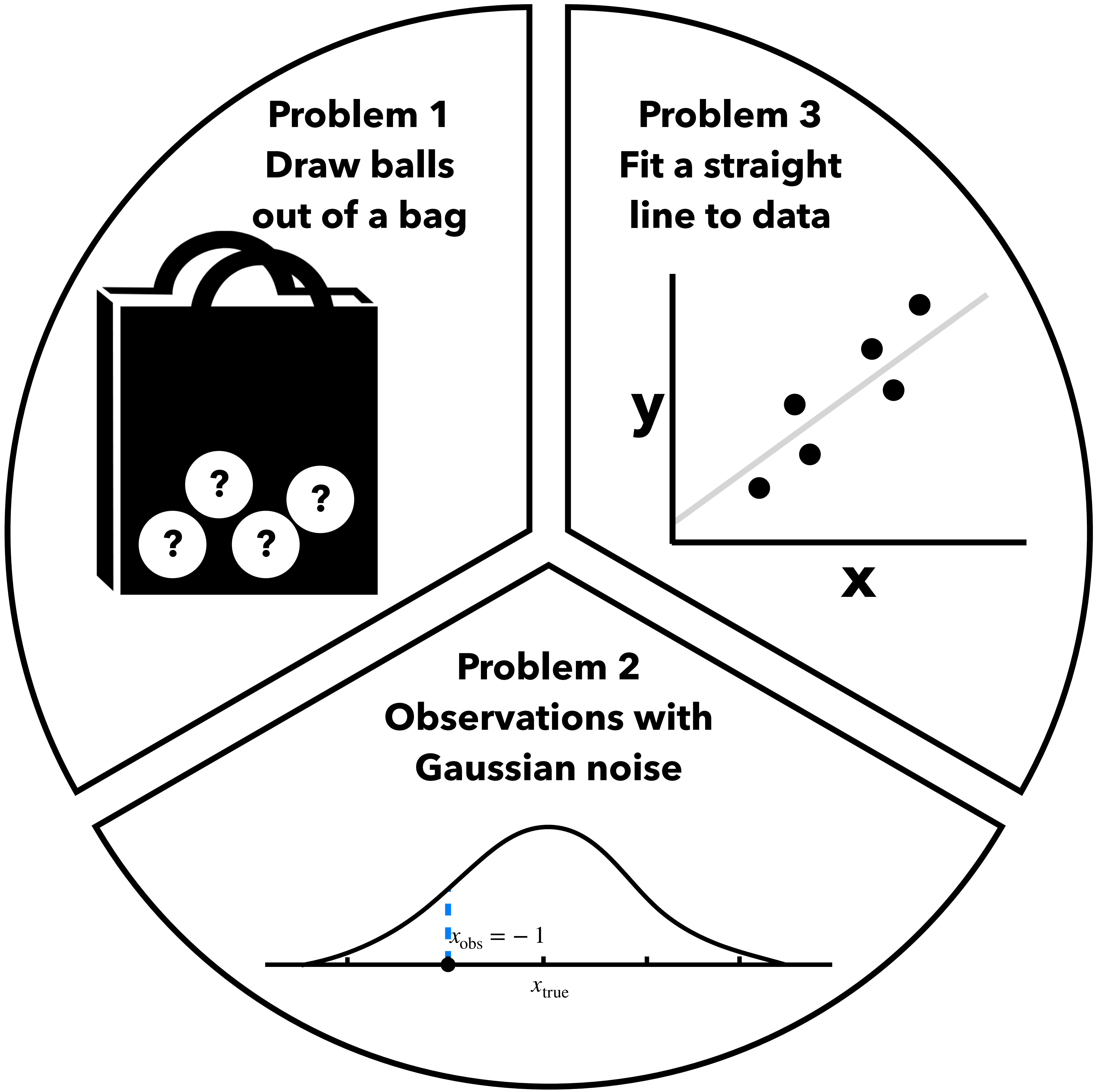
Instead of the “true” value of x, x_obs was observed.

Q: How do we infer x from x_obs?

$$x_{\text{obs}} = x_{\text{true}} + \mathcal{N}(0, \sigma^2)$$

$$\downarrow$$

$$x_{\text{obs}} \sim \mathcal{N}(x_{\text{true}}, \sigma^2)$$


Problem 1
Draw balls
out of a bag
?
?
?
?
Problem 3
Fit a straight
line to data
y
x
Problem 2
Observations with
Gaussian noise
$x_{\mathrm{obs}} = -1$
x_{true}

We have two parameters of interest, m (the slope of a line) and b (the intercept of a line).

We made an observation of y at some exact x value, but the observation was noisy.

Instead of the "true" value of y, y_obs was observed.

$$y_{\text{true}} = mx + b$$

$$y_{\text{obs}} \sim \mathcal{N}(y_{\text{true}}, \sigma^2)$$

! We are modeling with more variables than data points.

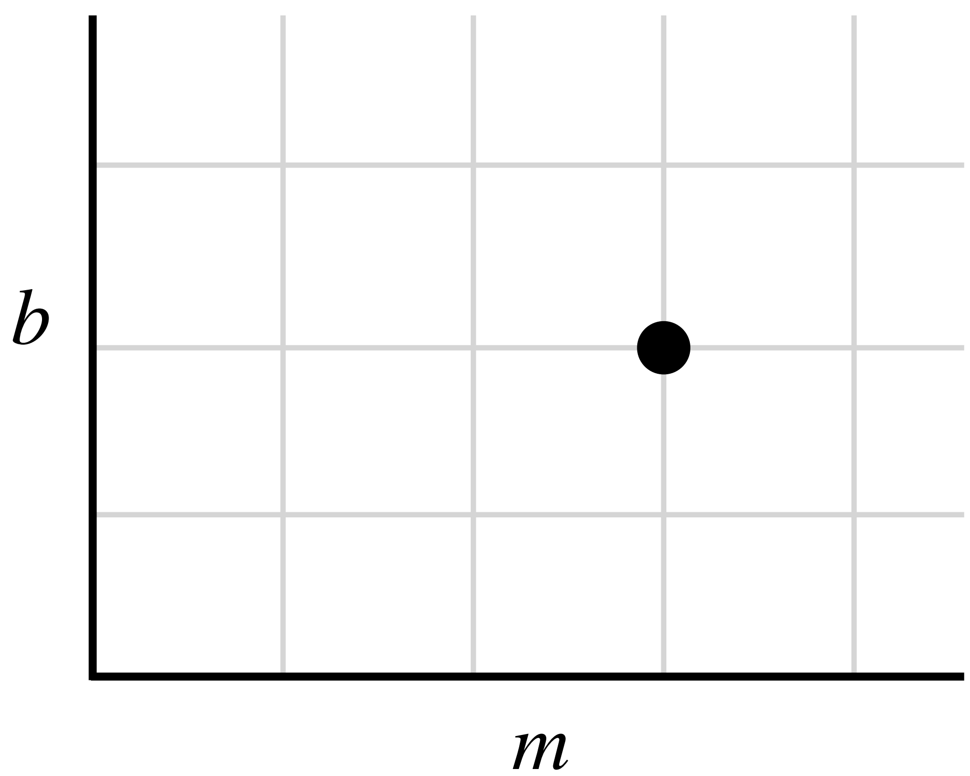
We have two parameters of interest, m (the slope of a line) and b (the intercept of a line).

We made an observation of y at some exact x value, but the observation was noisy.

Instead of the "true" value of y, y_obs was observed.

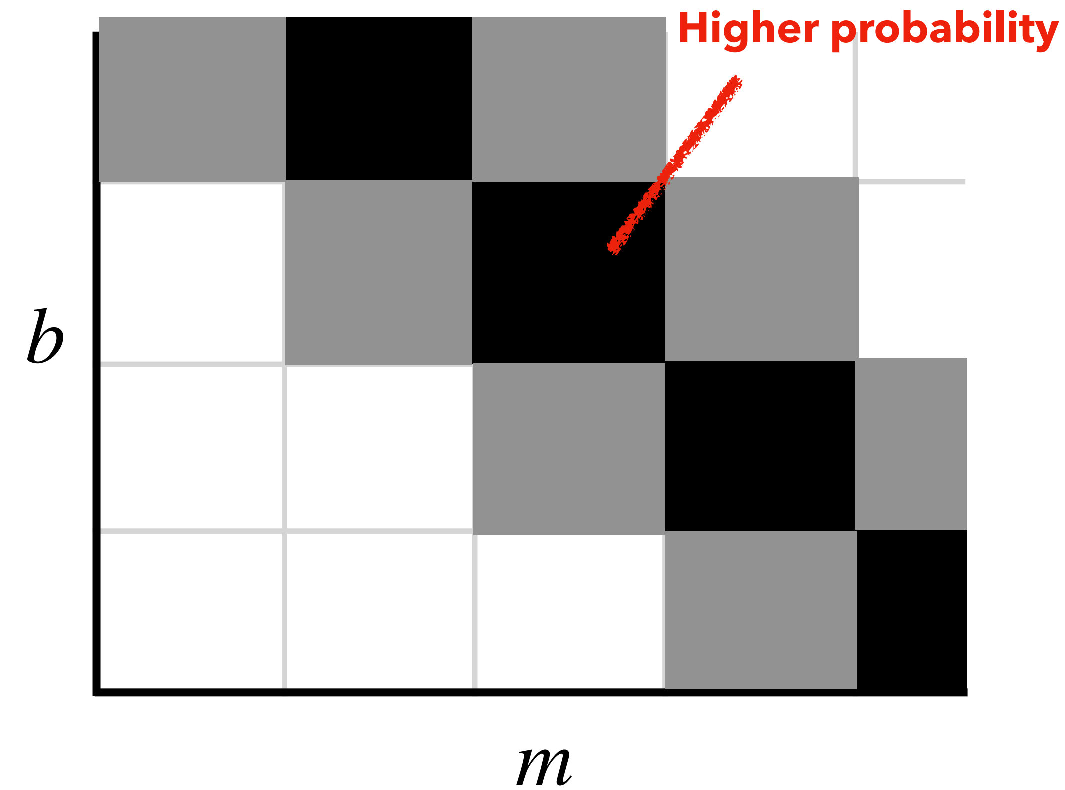
$$y_{\text{true}} = mx + b$$

$$y_{\text{obs}} \sim \mathcal{N}(y_{\text{true}}, \sigma^2)$$



For each pair of m, b, we calculate y_{true}, and the likelihood of y_{obs} given y_{true}.

The likelihood



We have two parameters of interest, m (the slope of a line) and b (the intercept of a line).

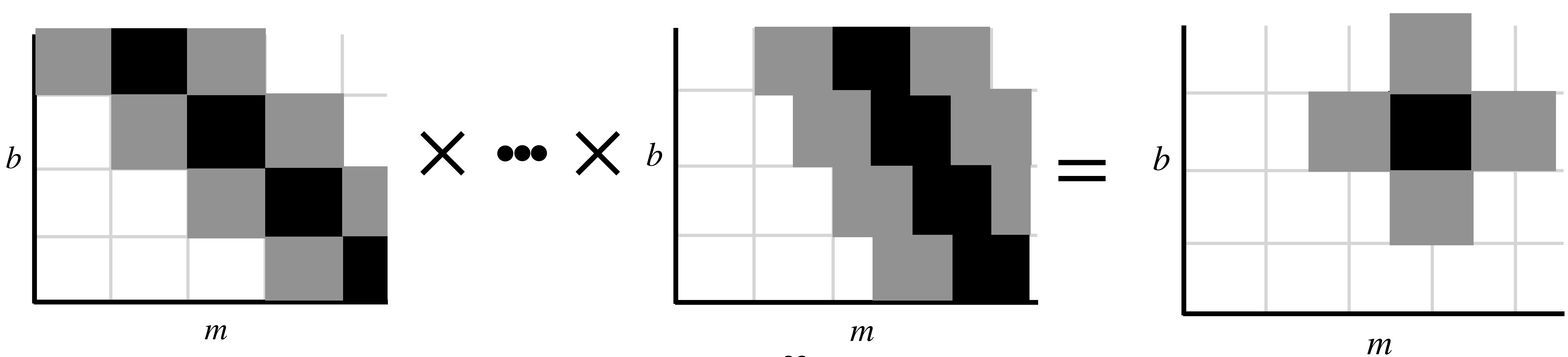
We made *N* observation of y at some exact x values, but the observation was noisy.

Instead of the "true" value of y, y_obs was observed.

$$y_i = mx_i + b$$

$$y_{\mathrm{obs,i}} \sim \mathcal{N}(y_i, \sigma^2)$$

For each y_i, we calculate its likelihood of $y_{\mathrm{obs,i}}$.



We have two parameters of interest, m (the slope of a line) and b (the intercept of a line).

We made *N* observation of y at some exact x values, but the observation was noisy.

Instead of the "true" value of y, y_obs was observed.

$$y_{\mathrm{i}} = m x_{\mathrm{i}} + b$$

$$y_{\mathrm{obs,i}} \sim \mathcal{N}(y_{\mathrm{i}}, \sigma^2)$$

$$p(\{y_{\mathrm{obs,i}}\} \mid \{y_{\mathrm{i}}\}, \sigma) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_{\mathrm{obs,i}} - y_i)^2}{2\sigma^2}\right]$$

$$\Big\downarrow \textbf{ natural log}$$

$$\log p(\{y_{\mathrm{obs,i}}\} \mid \{y_{\mathrm{i}}\}, \sigma) = -\frac{1}{2} \sum_{n=1}^{N} \left[\frac{(y_{\mathrm{obs,i}} - y_i)^2}{\sigma^2} + \log(2\pi\sigma^2)\right]$$

We have two parameters of interest, m (the slope of a line) and b (the intercept of a line).

We made *N* observation of y at some exact x values, but the observation was noisy.

Instead of the "true" value of y, y_obs was observed.

$$y_{\mathrm{i}} = m x_{\mathrm{i}} + b$$

$$y_{\mathrm{obs,i}} \sim \mathcal{N}(y_{\mathrm{i}}, \sigma^2)$$

$$m \sim \mathrm{Uniform}?$$

$$b \sim \mathrm{Uniform}?$$

! Uniform priors are not always uninformative priors.

See Brian's lecture tomorrow on "Priors, Likelihoods, Posteriors, and all that"

We have two parameters of interest, m (the slope of a line) and b (the intercept of a line).

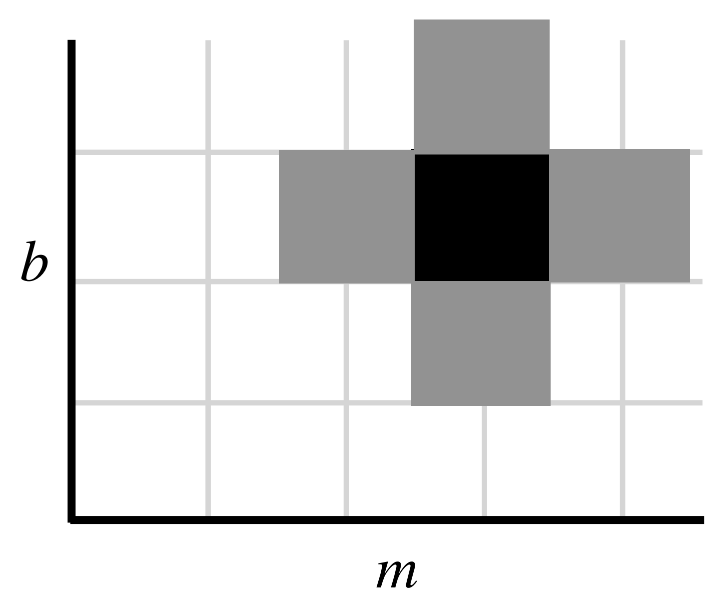
We made *N* observation of y at some exact x values, but the observation was noisy.

Instead of the "true" value of y, y_obs was observed.

$$y_\mathrm{i} = m x_\mathrm{i} + b$$

$$y_\mathrm{obs,i} \sim \mathcal{N}(y_\mathrm{i}, \sigma^2)$$

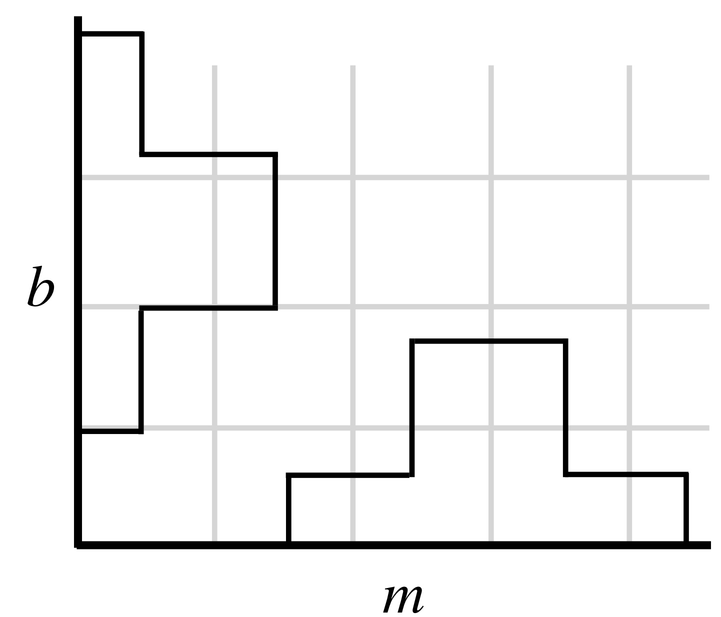
Joint posterior distribution $p(m, b \mid y_\mathrm{obs,i})$



To marginal posterior distribution $\longrightarrow$

$$p(m \mid y_\mathrm{obs,i}) = \int p(m, b \mid y_\mathrm{obs,i}) db$$

$$p(b \mid y_\mathrm{obs,i}) = \int p(m, b \mid y_\mathrm{obs,i}) dm$$



The approach we used in this lecture to compute the posterior is called "grid approximation".

In five steps,

1. **Build a grid for parameters of interest θ. The dimension of the grid depends on the number of parameters.**

2. **At each parameter value on the grid, calculate the prior $p(\theta_{\text{grid}})$.**

3. **At each parameter value on the grid, calculate the likelihood $p(D \mid \theta_{\text{grid}})$.**

4. **At each parameter value on the grid, multiply the likelihood by the prior $p(D \mid \theta_{\text{grid}})p(\theta_{\text{grid}})$.**

5. **Normalize the $p(D \mid \theta_{\text{grid}})p(\theta_{\text{grid}})$ by the sum of all values on the grid.**

References

- **McElreath, R. (2020). Statistical Rethinking: A Bayesian Course with Examples in R and Stan, 2nd Edition (2 ed.) CRC Press. (book)**