CSE 575
Statistical Machine Learning

Lecture 3

YooJung Choi

Fall 2022

Recap: Rules of probability

- Sum rule

$$p(X) = \sum_{Y} p(X, Y)$$

- Product rule

$$p(X, Y) = p(X|Y) \cdot p(Y)$$

Recap: Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y) \cdot p(Y)}{p(X)}$$

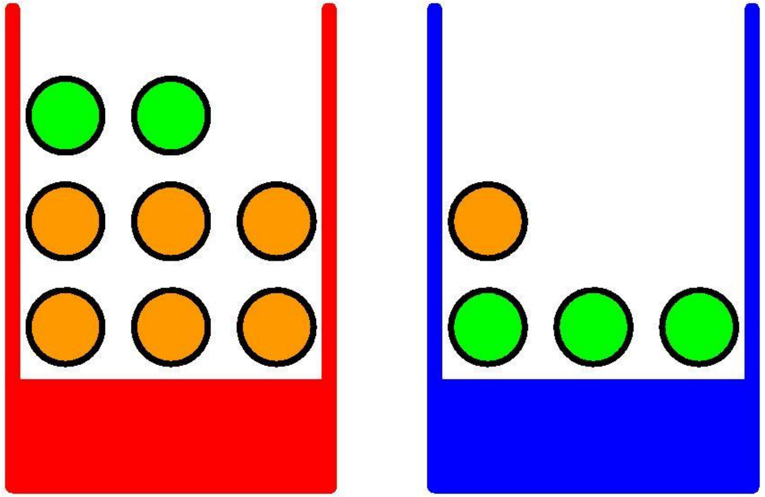
Sum rule

$$= \frac{p(X|Y) \cdot p(Y)}{\sum_Y p(X,Y)}$$

Product rule

$$= \frac{p(X|Y) \cdot p(Y)}{\sum_Y p(X|Y) \cdot p(Y)}$$

Bayes' Theorem interpretation



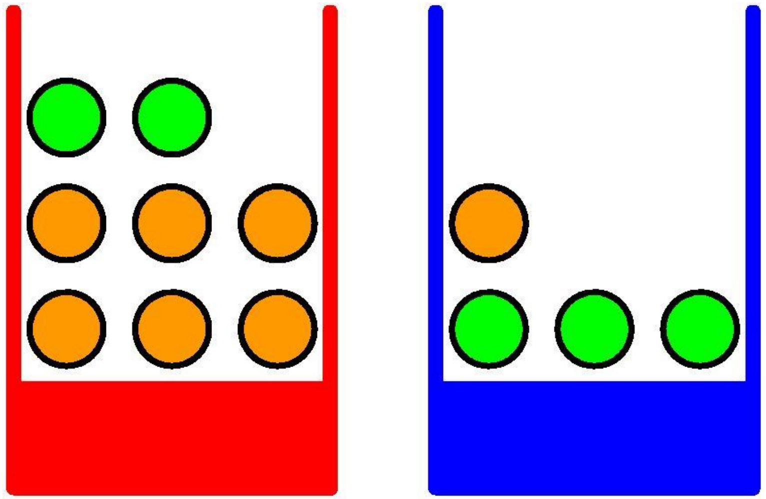
$$p(B|F) = \frac{p(F|B) \cdot p(B)}{p(F)}$$

Given: $p(B)$, $p(F|B)$

Q: $p(B|F)$?

- $p(B)$: "prior probability"
- $p(B|F)$: "posterior probability"
- $p(F)$: "probability of evidence"
- $p(F|B)$: "likelihood"

Bayes' Theorem interpretation



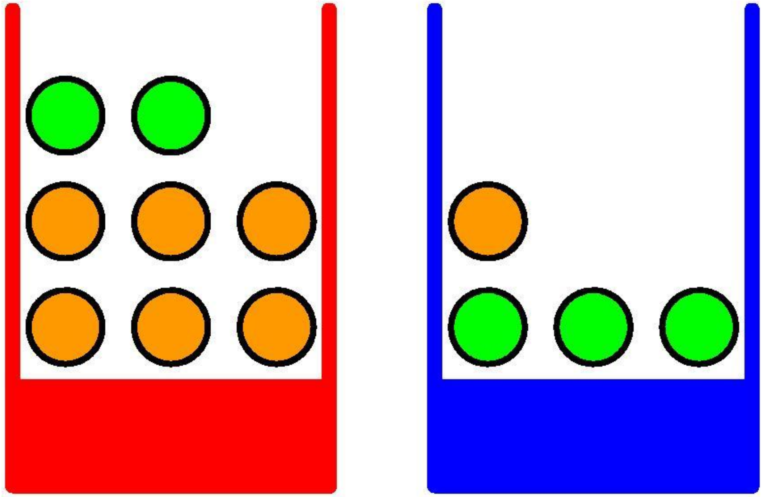
$$p(B|F) = \frac{p(F|B) \cdot p(B)}{p(F)}$$

$$Posterior = \frac{Likelihood \cdot Prior}{Evidence}$$

Given: $p(B)$, $p(F|B)$

Q: $p(B|F)$?

Bayes' Theorem interpretation



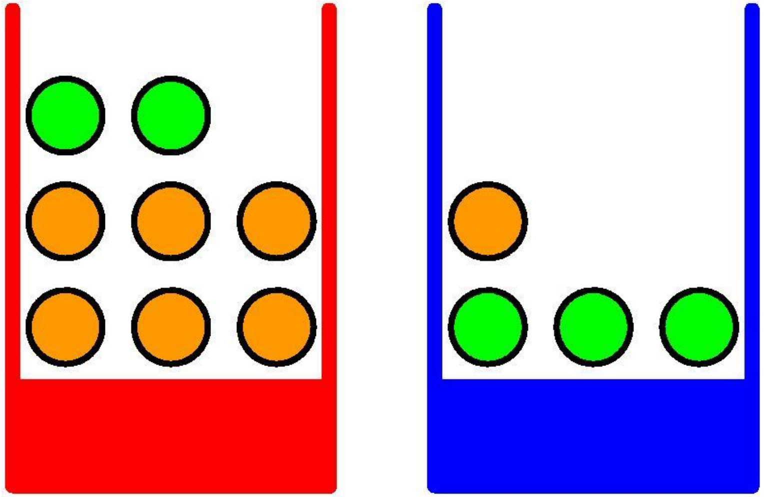
Given: $p(B)$, $p(F|B)$

Q: $p(B|F)$?

$$Posterior \propto Likelihood \cdot Prior$$

- Prior probability $p(B = b) = \frac{6}{10}$
- Given evidence $F = o$, the posterior probability *decreases* to
$$p(B = b|F = o) = \frac{1}{3}$$
- Intuition: the likelihood
$p(F = o|B = b) = 1/4$ is small

Bayes' Theorem interpretation



Given: $p(B)$, $p(F|B)$

Q: $p(B|F)$?

$$Posterior \propto Likelihood \cdot Prior$$

- Prior probability $p(B = r) = \frac{4}{10}$
- Given evidence $F = o$, the posterior probability *increases* to

$$p(B = r|F = o) = \frac{2}{3}$$

- Intuition: the likelihood

$p(F = o|B = r) = 3/4$ is large

Monty Hall Problem (video)

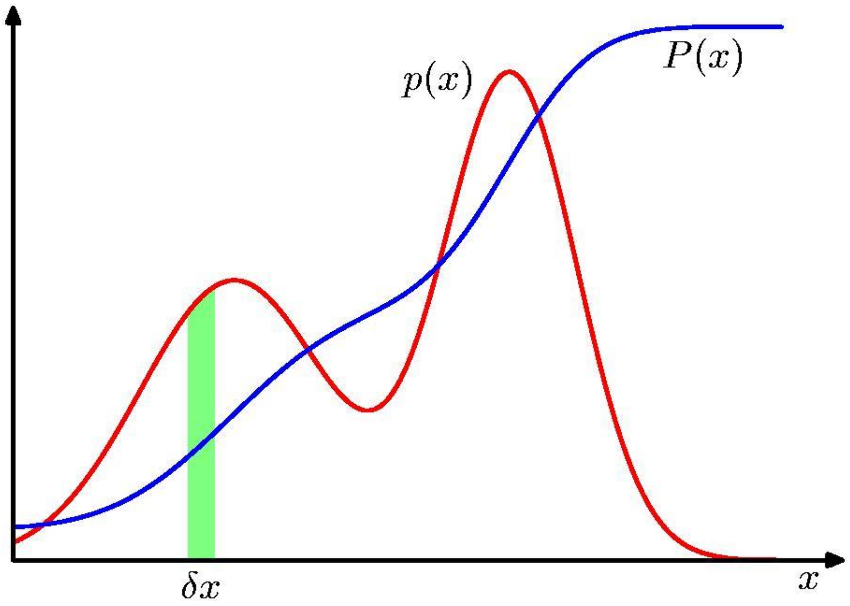
Independence

$$X \perp Y \quad \Leftrightarrow \quad p(X, Y) = p(X) \cdot p(Y) \quad \Leftrightarrow \quad p(Y|X) = p(Y)$$

Using product rule

- "The posterior probability is equal to the prior probability"
- i.e. Evidence X does not add any new information about Y

Probability densities



$$p(x \in (a, b)) = \int_a^b p(x)\,\mathrm{d}x$$

- Probability density function (PDF):

$$p(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$$

- Cumulative distribution function (CDF):

$$P(z) = \int_{-\infty}^{z} p(x)\,\mathrm{d}x$$

Rules of probability still hold

- Sum rule

$$p(x) = \int p(x, y)\, \mathrm{d}y$$

- Product rule

$$p(x, y) = p(y|x) \cdot p(x)$$

- Bayes' theorem

$$p(y|x) = \frac{p(x|y) \cdot p(y)}{p(x)}$$

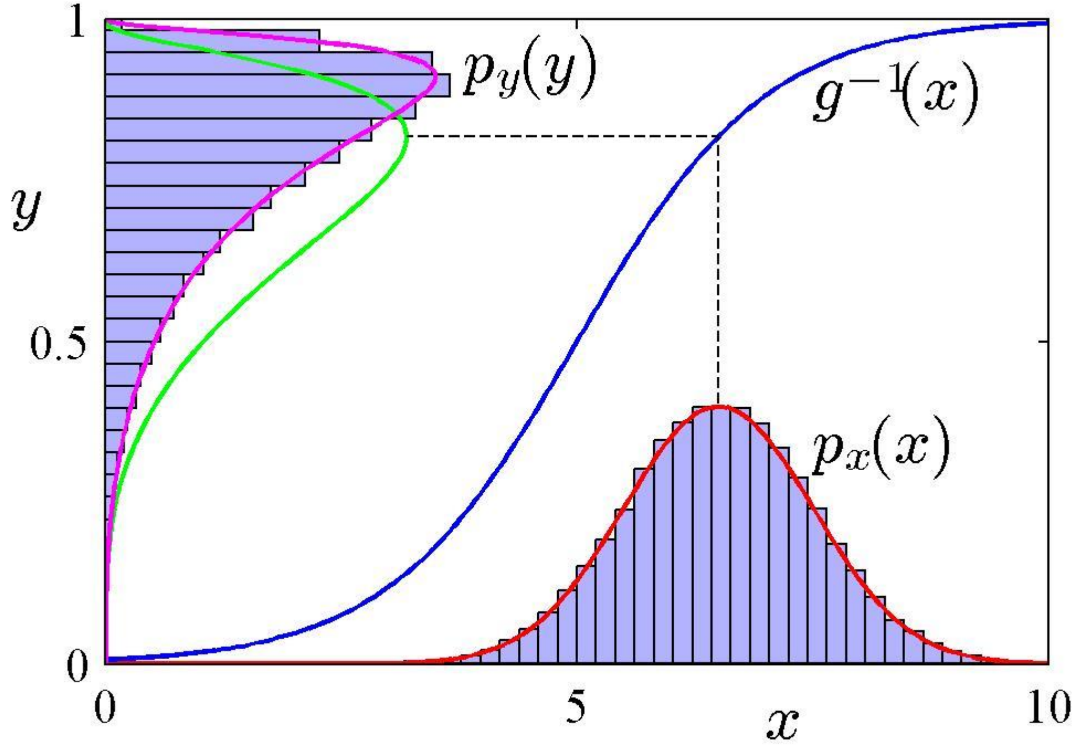
Transformed densities

- Consider a non-linear change of variable $x = g(y)$
- A simple function $f(x)$ becomes $\tilde{f}(g(y))$
- A density function $p_x(x)$ becomes $p_y(y)$ such that:

$$p_x(x)\delta x \simeq p_y(y)\delta y$$

$$\begin{aligned} \Rightarrow \quad p_y(y) &= p_x(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \\ &= p_x(g(y))\,|g'(y)| \end{aligned}$$

Transformed densities



[Guoliang Xue]

Expectation

- Expectation: average value of $f(x)$ under distribution $p(x)$

$$\mathbb{E}[f] = \sum_x p(x) f(x) \qquad\qquad \mathbb{E}[f] = \int p(x) f(x)\,\mathrm{d}x.$$

- Can be approximated as: $\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^{N} f(x_n).$

- Expectation over one variable: $\mathbb{E}_x[f(x, y)]$

- Conditional expectation: $\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$

Variance

- Variance: how much does $f(x)$ vary around its mean

$$\mathrm{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right]$$

- Equivalently, $\mathrm{var}[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2.$

- Variance of the variable: $\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2.$

Covariance

- Covariance: how much do two variables vary together

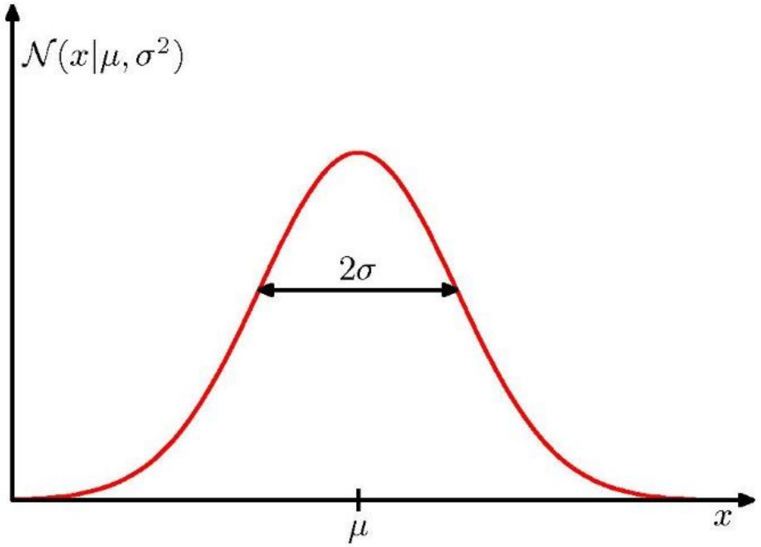
$$\begin{aligned}\mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y]\end{aligned}$$

The Gaussian distribution

- Normal / Gaussian distribution:

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

parameterized by mean μ, variance σ^2



- Is it a valid probability density?

$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$

The Gaussian distribution

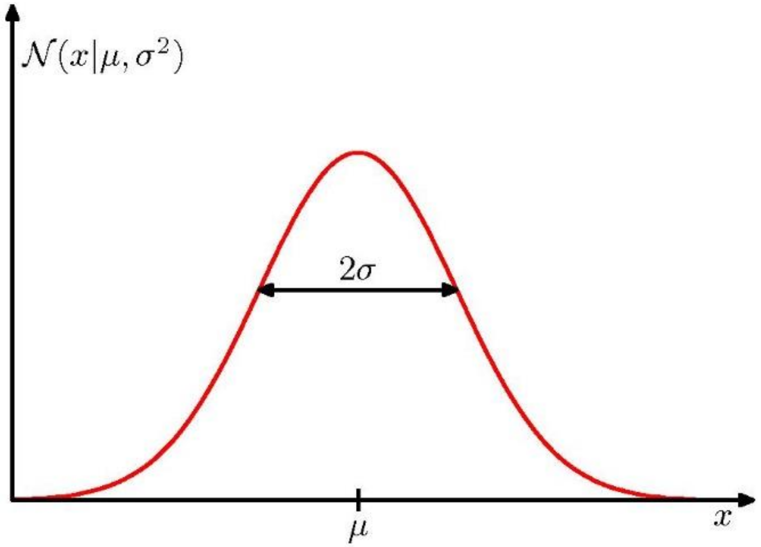
- Meanings of the parameters:
- The average value of a Gaussian:

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x \,\mathrm{d}x = \mu$$

- The second order moment:

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x^2 \,\mathrm{d}x = \mu^2 + \sigma^2$$

- Variance: $\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$



Multivariate Gaussian

- Gaussian distribution over a D-dimensional vector **x**:

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

- D-dimensional vector $\boldsymbol{\mu}$: mean
- D x D matrix $\boldsymbol{\Sigma}$: covariance ($\mathrm{cov}[\mathbf{x}] = \mathrm{cov}[\mathbf{x}, \mathbf{x}]$)

$$\begin{aligned} \mathrm{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}]. \end{aligned}$$