Probability & Statistics Overview

NEU 560
Jonathan Pillow

Fall 2020
Lecture 8
parameter $\theta$ \quad \rightarrow \quad P(x|\theta) \quad \rightarrow \quad X

model

samples

• “probability distribution”
• “events”
• “random variables”
parameter space

model

samples

• “probability distribution”
• “events”
• “random variables”

\( P(x | \theta) \)

\( \theta \)

• parameter

• samples

• parameter space

• sample space
parameters

samples

model

\[ P(x|\theta) \]

examples

1. coin flipping
   \[ \theta = p(\text{“heads”}) \]
   \[ X = \text{“H” or “T”} \]

2. spike counts
   \[ \theta = \text{mean spike rate} \]
   \[ X \in \{0, 1, \ldots\} \]
Probability vs. Statistics

Parameter space

\( \theta \)

Parameter

Model

\( P(x|\theta) \)

Probability

Samples

Coin flipping

\( \theta = 0.3 \)

T, T, H, T, H, T, T, T, T, T, ....
Probability vs. Statistics

Parameter space → Model \( P(x|\theta) \) → Sample space

Parameter

? \( \theta = ? \)

"inverse probability"

T, T, H, T, H, T,
T, T, T, H, T,
H, T, H, H, T, T
continuous probability distribution
takes values in a continuous space, e.g., \( x \in \mathbb{R} \)

probability density function (pdf):

\[
f(x)
\]

\[
x
\]

- \( P(x = a) = 0 \)
- \( P(a < x < b) = \int_{a}^{b} f(x) \, dx \)
- \( \int_{-\infty}^{\infty} f(x) \, dx = 1 \)
discrete probability distribution

takes finite (or countably infinite) number of values, eg \( x \in \mathbb{N} \)

**probability mass function (pmf):**

\[
f(x) \quad x
\]

- \( P(x = a) = f(a) \)
- \( \sum_{i=1}^{N} f(x_i) = 1 \)
some friendly neighborhood distributions

**Continuous**

Gaussian

\[ P(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right] \]

Multivariate Gaussian

\[ P(x_n; \mu, \Lambda) = \frac{1}{(2\pi)^{n/2} |\Lambda|^{1/2}} \exp \left[ -\frac{1}{2} (x - \mu)^T \Lambda^{-1} (x - \mu) \right] \]

Exponential

\[ P(x; a) = ae^{-ax} \]

**Discrete**

Bernoulli

\[ P(x; p) = p^x \cdot (1 - p)^{(1-x)} \]  
coin flipping

Binomial

\[ P(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \]  
sum of n coin flips

Poisson

\[ P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \]  
sum of n coin flips with P(heads)=\lambda/n, in limit n→∞
joint density

\[ P(x, y) \]

- positive
- sums to 1

\[ \iint P(x, y) \, dx \, dy = 1 \]
marginalization ("integration")

\[ P(x, y) \]

\[ P(x) = \int P(x, y) \, dy \]
marginalization ("integration")

\[ P(x, y) \]

\[ P(y) = \int P(x, y) \, dx \]
conditionalization ("slicing")

\[ P(x, y) \]

\[ P(y | x = -1) = \frac{P(y, x = -1)}{P(x = -1)} \]

("joint divided by marginal")
conditionalization ("slicing")

\[
P(x, y)
\]

\[
P(y|x = 1) = \frac{P(y, x = 1)}{P(x = 1)}
\]

("joint divided by marginal")
conditionalization ("slicing")

\[ P(x, y) \]

conditional

\[ P(y|x = 1) = \frac{P(y, x = 1)}{P(x = 1)} \]

marginal

\[ P(y) \]
conditional densities

\[ P(y|x) \]

\[ P(y|x) = \frac{P(x,y)}{P(x)} \]
conditional densities

\[ P(x|y) \]

\[ P(x|y) = \frac{P(x, y)}{P(y)} \]
Bayes’ Rule

\[ P(y|x) = \frac{P(x,y)}{P(x)} \quad \quad P(x|y) = \frac{P(x,y)}{P(y)} \]

\[ P(x,y) = P(y|x)P(x) = P(x|y)P(y) \]

Bayes’ Rule

\[ P(x|y) = \frac{P(y|x)P(x)}{P(y)} \]

conditional densities
likelihood
prior
posterior
marginal probability of y ("normalizer")
Expectations ("averages")

Expectation is the weighted average of a function (of a random variable) according to the distribution (of that random variable)

\[
\mathbb{E}[f(x)] = \sum_i f(x_i)P(x_i)
\]

or

\[
\mathbb{E}[f(x)] = \int f(x)P(x)dx
\]

Corresponds to taking weighted average of \(f(X)\), weighted by how probable they are under \(P(x)\).
Expectations ("averages")

Expectation is the weighted average of a function (of a random variable) according to the distribution (of that random variable)

**discrete**

\[
\mathbb{E}[f(x)] = \sum_i f(x_i)P(x_i)
\]

**continuous**

\[
\mathbb{E}[f(x)] = \int f(x)P(x)dx
\]

Monte Carlo evaluation of an expectation:

1. draw samples from distribution: \( x^{(i)} \sim P(x) \) for \( i = 1 \) to \( N \)

2. average

\[
\mathbb{E}[f(x)] \approx \frac{1}{N} \sum_{i=1}^{N} f(x^{(i)})
\]
Expectations ("averages")

Expectation is the weighted average of a function (of a random variable) according to the distribution (of that random variable)

\[ \mathbb{E}[f(x)] = \sum_i f(x_i) P(x_i) \quad \text{discrete} \]
\[ \mathbb{E}[f(x)] = \int f(x) P(x) dx \quad \text{continuous} \]

It's really just a dot product!

\[ \mathbb{E}[f(x)] = \vec{P} \cdot \vec{f} \quad \vec{P} = \begin{bmatrix} P(x_1) \\ \vdots \\ P(x_m) \end{bmatrix} \quad \vec{f} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{bmatrix} \]

Thus, expectation is a linear function:

\[ \mathbb{E}[af(x) + b] = a\mathbb{E}[f(x)] + b \]
Expectations ("averages")

The two most important expectations (also known as "moments"):

• Mean: $E[x]$ (average value of RV)
• Variance: $E[(x - E[x])^2]$ (average squared dist between $X$ and its mean).

Note: expectations don’t always exist!

e.g. Cauchy: $P(x) = \frac{1}{\pi(1 + x^2)}$ has no mean!
independence

$$P(x, y)$$

$$P(y)$$

$$P(x)$$
Definition: $x, y$ are independent iff

$$P(x, y) = P(x)P(y)$$
**independence**

**Definition:** $x, y$ are independent iff

$$P(x, y) = P(x)P(y)$$

In linear algebra terms:

$$P(x, y) = \vec{p}_y \vec{p}_x^T$$

(outer product)
Definition: \( x, y \) are independent iff

\[ P(x, y) = P(x)P(y) \]

Alternative definition:

\[ P(y|x) = P(y) \]

All conditionals are the same!
independence

Definition: $x, y$ are independent iff

$$P(x, y) = P(x)P(y)$$

Alternative definition:

$$P(y|x) = P(y)$$

All conditionals are the same!
Correlation vs. Dependence

1. Correlation

\[
\text{corr}(x, y) = \frac{\mathbb{E}[(x - \bar{x})(y - \bar{y})]}{\sqrt{\text{var}(x)\text{var}(y)}}
\]

Mean of \(y|x\) changes systematically with \(x\)

positive correlation

negative correlation
Correlation vs. Dependence

1. Correlation

\[ \text{corr}(x, y) = \frac{\mathbb{E}[(x - \bar{x})(y - \bar{y})]}{\sqrt{\text{var}(x)\text{var}(y)}} \]

2. Dependence

- arises whenever \( P(x, y) \neq P(x)P(y) \)
- can be quantified by mutual information: \( \text{MI}(x, y) = D_{KL}(P(x, y), P(x)P(y)) \)
- \( \text{MI}=0 \Rightarrow \text{independence} \)
Correlation vs. Dependence

**Q:** Can you draw a distribution that is *uncorrelated* but *dependent*?
Correlation vs. Dependence

Q: Can you draw a distribution that is *uncorrelated* but *dependent*?

“Bowtie” dependencies in natural scenes:

(uncorrelated but dependent)

\[
P(\text{filter 2 output} \mid \text{filter 1 output})
\]

Flower image: [Schwartz & Simoncelli 2001]
Is this distribution independent?

\[ P(x, y) \]
Is this distribution independent?

\[ P(y|x) \]
Is this distribution independent?

No! Conditionals over y are different for different x!
FUN FACT:

Independent Gaussian is the only distribution that is both:

- independent (equal to the product of its marginals)
- spherically symmetric: \( P(\vec{x}) = P(U\vec{x}) \)

**Corollary:** circular scatter / contour plot not sufficient to show independence!
Summary

• continuous & discrete distributions
• marginalization (splatting)
• conditionalization (slicing)
• Bayes’ rule (prior, likelihood, posterior)
• Expectations
• Independence & Correlation