Probabilistic Machine Learning

Tutorial 2

Start: 10.15

Keep your solution open in the background

Alexandra Gessner

27 April 2020

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Faculty of Science
Department of Computer Science
Chair for the Methods of Machine Learning

Tutorials over Zoom

quick reminder

New: Co-tutor present to assist tutor (chat etc.)

Communication & interaction within the zoom tutorial (controlled by tutors)

- **Speak:** Participants → raise hand → unmute

Exercise submissions

- Stick to the naming scheme, please: `02_<YourSurname>_<YourMatrikelnummer>.pdf`
- Make sure **figures** are visible in your submission

Shifting the tutorial?

- There will be a poll: Monday vs. Thursday

EXAMple: Beta-Bernoulli

Given:

N binary observations $X := [x_1, \ldots, x_N]$ iid. from Bernoulli distribution (likelihood of f)

$$p(x_i \,|\, f) = f^{\,x_i} \cdot (1-f)^{1-x_i} \qquad \text{for } i = 1, \ldots, N$$

Beta prior on f

$$p(f \,|\, a, b) = \mathcal{B}(f; a, b) := \frac{1}{B(a,b)} f^{\,a-1} \cdot (1-f)^{b-1} \qquad \text{with } a, b > 0, \quad f \in [0,1].$$

Question: What is the posterior $p(f \,|\, X)$?

Side note on the likelihood:

Never say *"the likelihood of the data"*. Always say *"the likelihood of the parameters"*. The likelihood function is not a probability distribution [in the parameters].
(David J. C. MacKay: Information Theory, Inference, and Learning Algorithms)

EXAMple

Conjugate priors

$$p(x_i \mid f) = f^{\,x_i} \cdot (1-f)^{1-x_i} \qquad \text{(L)}$$

$$p(f \mid a, b) \propto f^{\,a-1} \cdot (1-f)^{b-1} \qquad \text{(P)}$$

$$\begin{aligned}
p(f \mid X) = \frac{p(X \mid f)p(f)}{p(X)} &\propto p(X \mid f)p(f) && \text{Bayes' theorem}\\
&\overset{\text{iid.}}{=} p(f) \prod_{i=1}^{N} p(x_i \mid f)\\
&\propto f^{\,a-1} \cdot (1-f)^{b-1} \prod_{i=1}^{N} f^{\,x_i} \cdot (1-f)^{1-x_i} && \text{plug in (L),(P)}\\
&= f^{a-1}(1-f)^{b-1} \cdot f^{\sum_{i=1}^{N} x_i} (1-f)^{\sum_{i=1}^{N} 1-x_i}\\
&= f^{a+\sum_{i=1}^{N} x_i - 1}\ (1-f)^{b+N-\sum_{i=1}^{N} x_i - 1}\\
&= f^{\tilde{a}-1}\ (1-f)^{\tilde{b}-1}\\
&\propto \mathcal{B}(f; \tilde{a}, \tilde{b})
\end{aligned}$$

Theory question: Transforming random variables

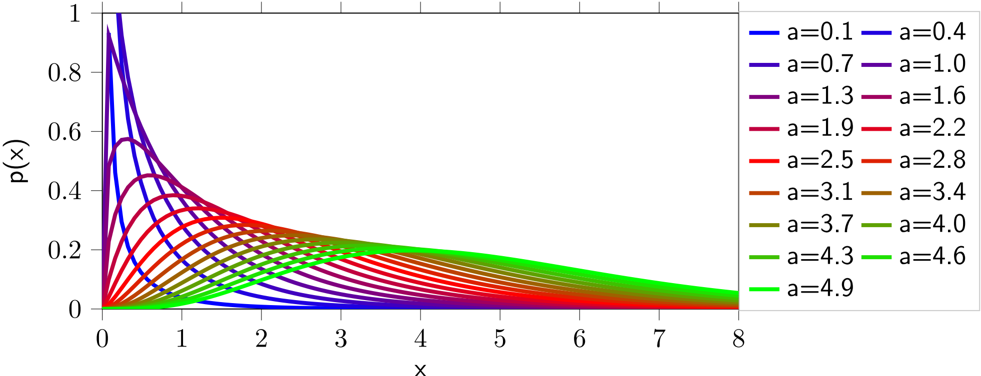
Theory Question: A Beta RV from two Gamma RVs

The Gamma distribution

Task: Given $X \sim \mathcal{G}(a, 1)$ and $Y \sim \mathcal{G}(b, 1)$, show that $Z = \frac{X}{X+Y} \sim \mathcal{B}(a, b)$

$$\mathcal{G}(\xi; \alpha, \beta) = \frac{\beta^\alpha \cdot \xi^{\alpha-1} e^{-\beta\xi}}{\Gamma(\alpha)}$$

Watch out: Two definitions of the Gamma distribution out there: $\beta = \frac{1}{\theta}$ (cf. lecture 4, slide 15)



Theory Question: (a) A Beta RV from two Gamma RVs

Task: Given $X \sim \mathcal{G}(a, 1)$ and $Y \sim \mathcal{G}(b, 1)$, show that $Z = \frac{X}{X+Y} \sim \mathcal{B}(a, b)$

Reminder from lecture:

Theorem (Transformation Law, general)

Let $X = (X_1, \ldots, X_d)$ have a joint density p_X. Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be continously differentiable and injective, with non-vanishing Jacobian J_g. Then $Y = g(X)$ has density

$$p_Y(y) = \begin{cases} p_X(g^{-1}(y)) \cdot |J_{g^{-1}}(y)| & \textit{if } y \textit{ is in the range of } g, \\ 0 & \textit{otherwise.} \end{cases}$$

$g(x, y) = \frac{x}{x+y}$ maps from $\mathbb{R}^2_+ \mapsto [0, 1]$, so we need a dummy variable to avoid a singular Jacobian, e.g. $g(x, y) = \left(x, \frac{x}{x+y}\right)$ for $x \neq 0$, or $g(x, y) = \left(x + y, \frac{x}{x+y}\right)$

Theory Question: (a) A Beta RV from two Gamma RVs

Solution: The function $g : \mathbb{R}_+^2 \mapsto \mathbb{R}_+ \times [0,1]$ as

$$g(x,y) = \left(x+y, \frac{x}{x+y}\right)$$

is injective. Thus, define the auxiliary variable $W = X + Y$. Then $X = ZW$ and $Y = W - ZW$.

Approach:

1. Find transformation $p_{X,Y}(x,y) \to p_{W,Z}(w,z)$
2. Marginalize $p_Z(z) = \int p_{W,Z}(w,z)\, dw$

First, compute Jacobian for the transform $p_{X,Y}(x,y) \to p_{W,Z}(w,z)$

$$\left|\frac{dg^{-1}(z,w)}{d(z,w)}\right| = \begin{vmatrix} \frac{dx}{dz} & \frac{dx}{dw} \\ \frac{dy}{dz} & \frac{dy}{dw} \end{vmatrix} = \begin{vmatrix} w & z \\ -w & 1-z \end{vmatrix} = w(1-z) + wz = w$$

Theory Question: (a) A Beta RV from two Gamma RVs

Memo: $X = ZW$ and $Y = W - ZW$; with Jacobian $\left|\frac{dg^{-1}(z,w)}{d(z,w)}\right| = w$

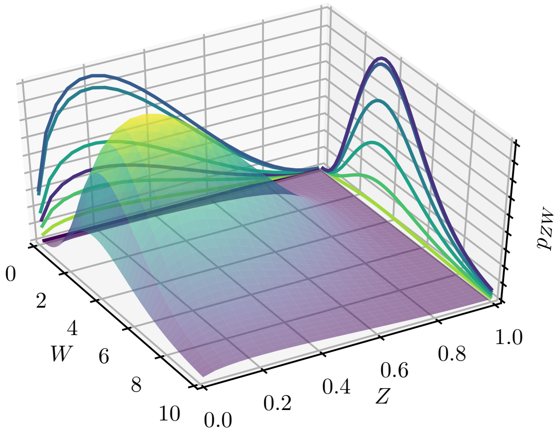
Next, compute the joint over Z and W is

$$\begin{aligned}
p_{Z,W}(z,w) &= p_{X,Y}(x(z,w), y(z,w)) \cdot \left|\frac{dg^{-1}(z,w)}{d(z,w)}\right| \\
&= p_X(x(z,w))\ p_Y(y(z,w)) \cdot w \\
&= \mathcal{G}(zw; a, 1)\ \mathcal{G}(w - zw; b, 1) \cdot w \\
&= \frac{1}{\Gamma(a)\Gamma(b)}(zw)^{a-1}e^{-zw}(w - zw)^{b-1}e^{-w+zw}w \\
&= \frac{1}{\Gamma(a)\Gamma(b)}w^{a+b-1}z^{a-1}(1-z)^{b-1}e^{-w} \\
&= p_Z(z)p_W(w)
\end{aligned}$$

Theory Question: (a) A Beta RV from two Gamma RVs

Joint over Z and W visualized

$$p_{Z,W}(z,w) \\ = p_Z(z)p_W(w) \\ \propto z^{a-1}(1-z)^{b-1}w^{a+b-1}e^{-w}$$



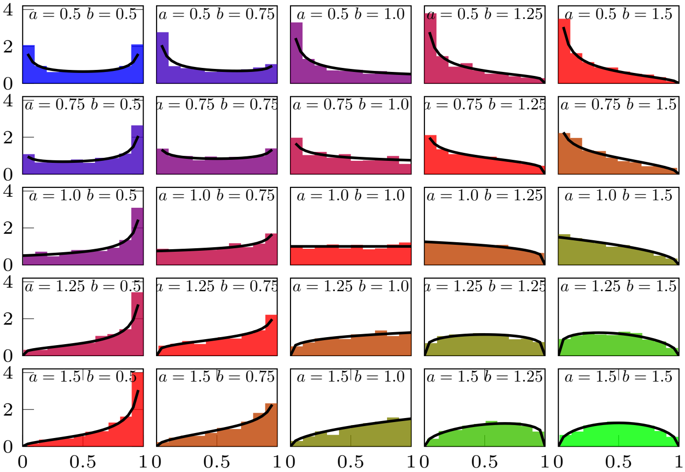
Theory Question: (a) A Beta RV from two Gamma RVs

Marginalize over W

$$
\begin{aligned}
p_Z(z) &= \int_0^\infty dw \; p_{Z,W}(z,w) \\
&= \frac{1}{\Gamma(a)\Gamma(b)} z^{a-1}(1-z)^{b-1} \int_0^\infty dw \; \underbrace{w^{a+b-1}e^{-w}}_{\propto \mathcal{G}(w;a+b,1)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} z^{a-1}(1-z)^{b-1} \\
&= \frac{1}{B(a,b)} z^{a-1}(1-z)^{b-1} \\
&= \mathcal{B}(z;a,b) \qquad \square
\end{aligned}
$$

Theory Question: (a) A Beta RV from two Gamma RVs

A Python example



$$X \sim \mathcal{G}(a, 1); \quad Y \sim \mathcal{G}(b, 1)$$
$$\Rightarrow Z = \frac{X}{X+Y} \sim \mathcal{B}(a, b)$$

```
import numpy as np
from scipy.stats import gamma

n_param = 5
N = 1000
a = np.linspace(0.5, 1.5, n_param)
b = np.linspace(0.5, 1.5, n_param)

for i in range(n_param):
    for j in range(n_param):
        x = gamma.rvs(a[i], size=N)
        y = gamma.rvs(b[j], size=N)
        z = x/(x+y)
```

Theory Question: (b) Mean of the Beta distribution

$$
\begin{aligned}
\mathbb{E}_{\mathcal{B}(z;a,b)}[z] &= \int_0^1 z\ \mathcal{B}(z;a,b)\,dz \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \underbrace{\int_0^1 z^a (1-z)^{b-1}\,dz}_{=B(a+1,b)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\cancel{\Gamma(b)}} \frac{\Gamma(a+1)\cancel{\Gamma(b)}}{\Gamma(a+b+1)} \quad \text{using } B(a+1,b) = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\
&= \frac{\cancel{\Gamma(a+b)}}{\cancel{\Gamma(a)}} \cdot \frac{a\cancel{\Gamma(a)}}{(a+b)\cancel{\Gamma(a+b)}} \quad \text{using } \Gamma(x+1) = x\Gamma(x) \\
&= \frac{a}{a+b}
\end{aligned}
$$

$\square$

Practical: Battleships I