Parametric statistical models

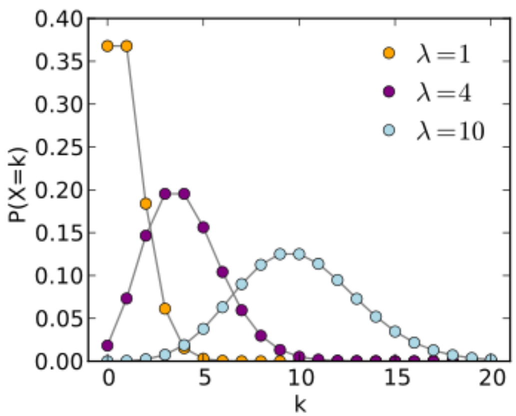
Parametric models

- A *(statistical) model* $\mathcal{P} = \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is a family of probability distributions indexed by a set Θ (the *parameter space*).
- In a *parametric (statistical) model*, the distributions are indexed by a finite number of parameters (i.e., $\Theta \subseteq \mathbb{R}^k$ for some $k < \infty$).
- Examples:
  - Bernoulli distributions on $\{0, 1\}$
  - Poisson distributions on $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$
  - Gaussian distributions on $\mathbb{R}$
  - Multivariate Gaussian distributions on $\mathbb{R}^d$

Poisson distributions $\mathrm{Poi}(\lambda)$ on $\mathbb{Z}_+$

- Indexed by positive scalar $\boldsymbol{\theta} = \lambda > 0$ called the *rate parameter*.
- The distribution P_λ is supported (i.e., non-zero) on $\mathbb{Z}_+$.

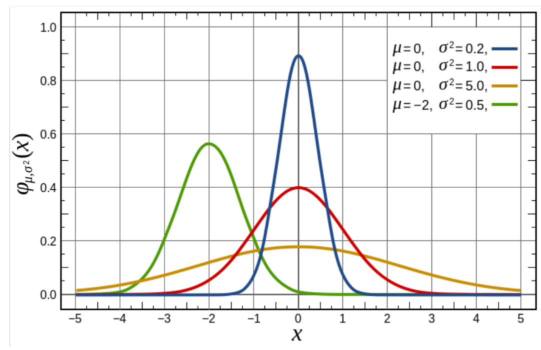
$$P_\lambda(x) = \frac{\exp(-\lambda)\lambda^x}{x!}.$$



Gaussian distributions $\mathrm{N}(\mu, \sigma^2)$ on $\mathbb{R}$

- Indexed by $\boldsymbol{\theta} = (\mu, \sigma^2)$: $\mu \in \mathbb{R}$ is the *mean*, $\sigma^2 > 0$ is the *variance*.
- The distribution P_{μ,σ^2} has the probability density

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$



Model fitting (parameter estimation)

- Suppose we have an observation x, and a parametric model $\mathcal{P}$ with parameter space Θ.
- **Model fitting (parameter estimation)**:
  pick some $P_{\boldsymbol{\theta}} \in \mathcal{P}$ (i.e., some $\boldsymbol{\theta} \in \Theta$) that best "fits" the observation x.
- A particular method: **maximum likelihood estimator (MLE)**
  - *Likelihood* of $\boldsymbol{\theta}$ given observation x:
  $$\mathcal{L}(\boldsymbol{\theta}; x) \ := \ P_{\boldsymbol{\theta}}(x)$$
  (replace $P_{\boldsymbol{\theta}}$ with its density if it has one).
  - Choose $\boldsymbol{\theta} \in \Theta$ with the highest likelihood.
  - Equivalent to choosing $\boldsymbol{\theta} \in \Theta$ with the highest *log-likelihood* $\log \mathcal{L}(\boldsymbol{\theta}; x)$.

Example: MLE for $\mathrm{Poi}(\lambda)$

- Suppose we observe $x = 4$.
  What is the MLE for the Poisson rate parameter λ?
- Log-likelihood of λ given $x = 4$:
  $$\log \mathcal{L}(\lambda; 4) \ = \ \log \frac{\exp(-\lambda)\lambda^4}{4!} \ = \ -\lambda + 4 \log \lambda - \log 4!$$
  which (by calculus) is maximized at $\lambda = 4$.
- Indeed, for any given observation $x \in \mathbb{Z}_+$, the MLE for λ is x.

Models for i.i.d. samples

- We often want a statistical model for an i.i.d. sample (say, of size n).
- **Technically**: $\mathcal{P}^n := \{P_{\boldsymbol{\theta}}^n : \boldsymbol{\theta} \in \Theta\}$, where $P_{\boldsymbol{\theta}}^n$ is the n-fold product of $P_{\boldsymbol{\theta}}^n$.

  If $P_{\boldsymbol{\theta}}$ is a probability distribution on $\mathcal{X}$, then $P_{\boldsymbol{\theta}}^n$ is the probability distribution on $\underbrace{\mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}}_{n \text{ times}}$, where
  $$P_{\boldsymbol{\theta}}^n(x_1, x_2, \ldots, x_n) \ = \ \prod_{i=1}^{n} P_{\boldsymbol{\theta}}(x_i) \, .$$
- Whenever we think of a data set (e.g., $\{x_i\}_{i=1}^n$) as an i.i.d. sample, we shall implicitly "lift" our statistical models $\mathcal{P}$ to their corresponding n-fold product form $\mathcal{P}^n$.

Example: MLE for $\mathrm{Poi}(\lambda)$ (again)

- Suppose we observe $\{x_i\}_{i=1}^n$ (regarded as an i.i.d. sample).
  What is the MLE for the Poisson rate parameter λ?
- Log-likelihood of λ given $\{x_i\}_{i=1}^n$:
  $$\log \mathcal{L}(\lambda; \{x_i\}_{i=1}^n) \ = \ \log \prod_{i=1}^{n} \frac{\exp(-\lambda)\lambda^{x_i}}{x_i!} \ = \ \sum_{i=1}^{n} -\lambda + x_i \log \lambda - \log x_i!$$
  which (by calculus) is maximized at $\lambda = \frac{x_1 + x_2 + \cdots + x_n}{n}$ (i.e., *sample mean*).

Example: MLE for $\mathrm{N}(\mu, \sigma^2)$

- Suppose we observe $\{x_i\}_{i=1}^n$ (regarded as an i.i.d. sample). What is the MLE for the Gaussian parameters (μ, σ^2)?
- Log-likelihood of (μ, σ^2) given $\{x_i\}_{i=1}^n$:

$$\log \mathcal{L}((\mu, \sigma^2); \{x_i\}_{i=1}^n) \;=\; \log \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right).$$

- Lo and behold, MLE for μ is the *sample mean*, and MLE for σ^2 is the *sample variance*.
- **Question**: What could go wrong?

Other approaches to parameter estimation

- MLE is a very powerful method, but not always the preferred choice.
- **Some other methods**:
  - Penalized maximum likelihood.
  - Method of moments.
  - Bayesian inference: don't select just a single $\boldsymbol{\theta} \in \Theta$; rather, choose an entire distribution over Θ (called the *posterior distribution*).

Key takeaways

1. Concept of a statistical model and the "lifting" to a model for an i.i.d. sample.
2. Derivation of maximum likelihood estimator for some simple models.