Advanced topics in Markov-chain Monte Carlo

Lecture 9:
Beyond sampling Part 1/2: Learning form finite sample sizes

Werner Krauth

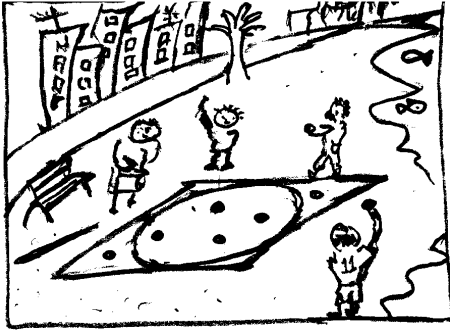
ICFP -Master Course Ecole Normale Supérieure, Paris, France

23 March 2022

References

- L. Wasserman ALL OF STATISTICS (Springer, 2004)
- W. Krauth STATISTICAL MECHANICS: ALGORITHMS AND COMPUTATIONS (Oxford, 2006)

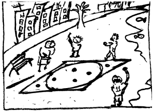
Frequentist interpretation of probabilities 1/3



- (Strong) law of large numbers:

$$P\left[\lim_{t\to\infty}\frac{1}{t}\sum_{t'=1}^{t}\mathcal{O}(t')=\mathbb{E}\left(\mathcal{O}\right)\right]=1 \qquad (1)$$

Frequentist interpretation of probabilities 2/3

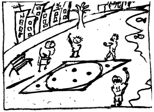


- Bernoulli distribution:

$$\xi_i = \begin{cases} 1 & \text{with prob } \theta \\ 0 & \text{with prob } 1-\theta \end{cases}$$

- $\theta = \pi/4$, but π unknown.
- Put (astuce!): $\eta_i = \xi_i - \theta$ with $\langle \eta_i \rangle = 0$ and $\text{Var}\,(\eta_i) = \text{Var}\,(\xi_i) = \theta(1-\theta) \leq 1/4$.

Frequentist interpretation of probabilities 3/3

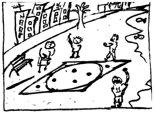


- η: RV with variance < $1/4$, and zero mean!!
- with 68% probability:

$$|\frac{1}{6}\sum_i \eta_i| < \sqrt{\frac{1}{4 \times 6 \times 0.32}} = 0.36$$

- $|\frac{1}{6}\sum_i \eta_i| = |\frac{4}{6} - \frac{\pi}{4}| < 0.36$. with 68% probability
- $\pi = \frac{16}{6} \pm 4 \times 0.36$.
- $\pi \in [1.23, 4.10]$ with 68% probability.

Bayesian interpretation of probabilities 1/3



- ξ_i: Bernoulli variable with unknown θ
- Outcome of run with $N = 6$: 4/6

Bayesian interpretation of probabilities 2/3

procedure naive-bayes-pi

1 $\pi_{\text{test}} \leftarrow$ ran $(0, 4)$ (sampled with the a priori probability)

$N_{\text{hits}} \leftarrow 0$

for $i = 1, \ldots,$ 6 **do**

$\quad$ $\{$ **if** (ran $(0,1) < \pi_{\text{test}}/4$) **then**

$\qquad$ $\{$ $N_{\text{hits}} \leftarrow N_{\text{hits}} + 1$

if ($N_{\text{hits}} \neq$ 4) **goto** 1 (reject π_{test})

output π_{test} (output with the a posteriori probability)

——

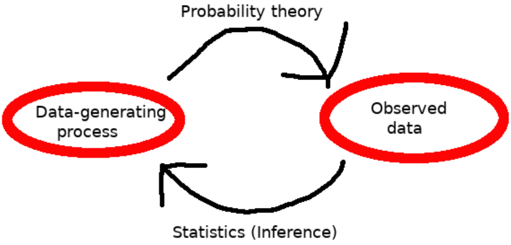
- Bernoulli trials with a certain a priori probability of π_{test} that obtain 4 successes with 6 trials.

Bayesian interpretation of probabilities 3/3

$$\int_0^4 \mathrm{d}\pi_{\text{test}}\ ,\quad \int_0^{16} \mathrm{d}(\pi_{\text{test}}^2)\ ,\quad \int_0^2 \mathrm{d}\sqrt{\pi_{\text{test}}}\ ,\ \ldots \underbrace{\int_{3\frac{10}{71}}^{3\frac{1}{7}} \mathrm{d}\pi_{\text{test}}}_{\text{Archimedes,}}\ ,\ \text{etc.}$$

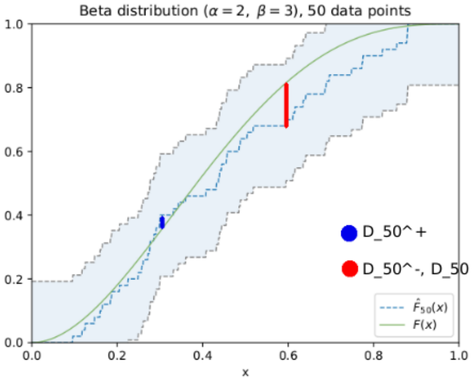
- Absence of neutral choice.

Non-parametric statistics 1/3



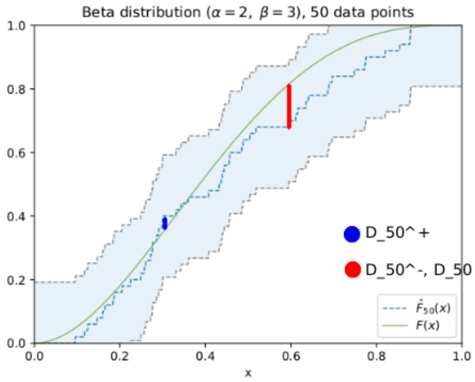
- Model (in statistics): A family of probability distributions.
- Parametric model: Family that can be described by a finite number of parameters
- Non-parametric model: any other ...
- NB: Children’s game is parametric ...

Non-parametric statistics 2/3



- Dvoretzky-Kieffer-Wolfowitz inequality
- A corridor around $\hat{F}_n$ (empirical CDF) fully contains CDF.

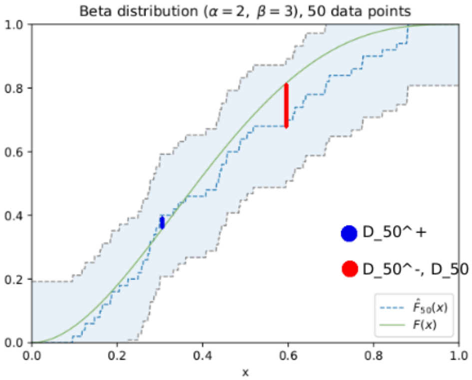
Non-parametric statistics 3/3



- Width of corridor $\propto 1/\sqrt{n}$:

$$\mathbb{P}\left[\sup_x |\hat{F}_n(x) - F(x)| > \epsilon\right] \leq 2\mathrm{e}^{-2n\epsilon^2}$$

Bootstrap 1/1



- Bootstrap: Replace $F(x)$ by $\hat{F}_n(x)$
- $\hat{F}_n(x)$: n samples from F
- Bootstrap: n samples from $\hat{F}_n$.