Introduction to Probability

Session 13: Example Class

Mateja Jamnik, Thomas Sauerwald

University of Cambridge, Department of Computer Science and Technology
email: {mateja.jamnik,thomas.sauerwald}@cl.cam.ac.uk

Easter 2023

UNIVERSITY OF CAMBRIDGE

Plan for Today

3 worked out examples:

1. Application of Central Limit Theorem
2. Bias and MSE of Estimators
3. Local Maxima ("Best-so-far Candidates") in the Secretary Problem

And plenty of time to answer your questions!

Example 1

Assume that an unknown fraction p of voters support a particular candidate. We poll $n = 100$ random voters and record by $\overline{X}_n := \frac{1}{n} \cdot (X_1 + X_2 + \cdots + X_n)$ the fraction of polled voters that support the candidate. Using the CLT, find an ϵ so that $\mathbf{P}\left[\left|\overline{X}_n - p\right| \leq \epsilon\right] \geq 0.95$.

Answer

- Clearly, $\mu = \mathbf{E}[X_i] = p$ and $\sigma^2 = \mathbf{V}[X_i] = p(1-p)$ are finite.
- We have

$$\mathbf{P}\left[\left|\overline{X}_n - p\right| \geq \epsilon\right] = \mathbf{P}\left[\overline{X}_n - p > \epsilon\right] + \mathbf{P}\left[\overline{X}_n - p \leq -\epsilon\right] \overset{!}{\leq} 0.05$$

- Remark: For simplicity (and as $n \geq 100$) we skip the continuity correction here
- $\overline{X}_n - p$ has already mean zero, only remains to scale it to get a r.v. with variance 1:

$$\mathbf{P}\left[\overline{X}_n - p \geq \epsilon\right] = \mathbf{P}\left[(\overline{X}_n - p) \cdot \sqrt{n}/\sigma \geq \frac{\sqrt{n} \cdot \epsilon}{\sigma}\right]$$

$$\overset{\text{(CLT)}}{\approx} 1 - \Phi\left(\frac{\sqrt{n} \cdot \epsilon}{\sigma}\right) \overset{!}{=} 0.025.$$

- Rearranging gives $\frac{\sqrt{n}\cdot\epsilon}{\sigma} = \Phi^{-1}(0.975) = 1.96$
- $\sigma = \sqrt{p(1-p)}$, but p is unknown $\rightsquigarrow$ assume σ is as large as possible, i.e., $\sigma = 1/2$:

$$\epsilon \geq 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{1}{20} \qquad \Rightarrow \qquad \epsilon \approx 0.098.$$

- We also have $\mathbf{P}\left[\overline{X}_n - p \leq -\epsilon\right] \approx \Phi\left(-\frac{\sqrt{n}\cdot\epsilon}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}\cdot\epsilon}{\sigma}\right)$, hence for the same choice of ϵ, we have $\mathbf{P}\left[\overline{X}_n - p \leq -\epsilon\right] \leq 0.025$.

Example 2 [source: Dekking et al., Exercise 20.3]

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. samples with distribution $Exp(\lambda)$. We would like to estimate the unknown mean $1/\lambda$. Let $T_1 := \overline{X}_n = \frac{1}{n} \cdot (X_1 + X_2 + \ldots + X_n)$ be the sample mean.

1. Define $M_n := \min(X_1, X_2, \ldots, X_n)$. What is the distribution of M_n?
2. Find an unbiased estimator T_2 for $1/\lambda$ based on M_n.
3. Which of the two estimators T_1 or T_2 is preferable?

Answer

1. We have for $x \geq 0$,

$$\begin{aligned} \mathbf{P}[M_n \geq x] &= \mathbf{P}\left[\cap_{i=1}^{n}(X_i \geq x)\right] \\ &= \prod_{i=1}^{n} \mathbf{P}[X_i \geq x] \\ &= (\mathbf{P}[X_1 \geq x])^n \\ &= \left(e^{-\lambda \cdot x}\right)^n \\ &= e^{-(\lambda \cdot n) \cdot x}. \end{aligned}$$

Hence $M_n \sim Exp(\lambda \cdot n)$. Thus $\mathbf{E}[M_n] = 1/(\lambda \cdot n)$.

Example 2 [source: Dekking et al., Exercise 20.3]

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. samples with distribution $Exp(\lambda)$. We would like to estimate the unknown mean $1/\lambda$. Let $T_1 := \overline{X}_n = \frac{1}{n} \cdot (X_1 + X_2 + \ldots + X_n)$ be the sample mean.

1. Define $M_n := \min(X_1, X_2, \ldots, X_n)$. What is the distribution of M_n?
2. Find an unbiased estimator T_2 for $1/\lambda$ based on M_n.
3. Which of the two estimators T_1 or T_2 is preferable?

Answer

2. Recall $\mathbf{E}[M_n] = 1/(\lambda \cdot n)$. Hence an unbiased estimator for $1/\lambda$ is:

$$T_2 := n \cdot M_n.$$

Example 2 [source: Dekking et al., Exercise 20.3]

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. samples with distribution $Exp(\lambda)$. We would like to estimate the unknown mean $1/\lambda$. Let $T_1 := \overline{X}_n = \frac{1}{n} \cdot (X_1 + X_2 + \ldots + X_n)$ be the sample mean.

1. Define $M_n := \min(X_1, X_2, \ldots, X_n)$. What is the distribution of M_n?
2. Find an unbiased estimator T_2 for $1/\lambda$ based on M_n.
3. Which of the two estimators T_1 or T_2 is preferable?

Answer

2. Recall $\mathbf{E}[M_n] = 1/(\lambda \cdot n)$. Hence an unbiased estimator for $1/\lambda$ is:

$$T_2 := n \cdot M_n.$$

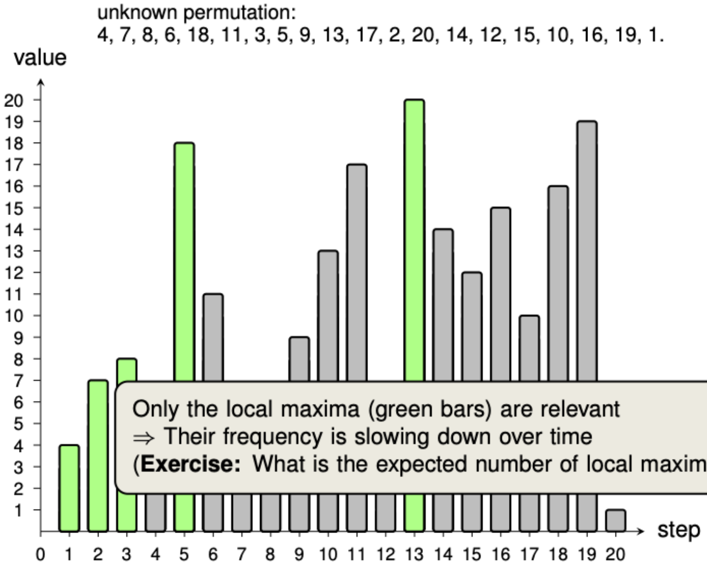
3. Both T_1 and T_2 are unbiased, therefore by the bias-variance decomposition:

$$\mathbf{MSE}[T_1] = \underbrace{\left(\mathbf{E}[T_1] - \frac{1}{\lambda}\right)^2}_{=0} + \mathbf{V}[T_1] = \mathbf{V}[T_1] = \frac{1}{n^2} \cdot (n \cdot \mathbf{V}[X_1]) = \frac{1}{n} \cdot \frac{1}{\lambda^2}$$

$$\mathbf{MSE}[T_2] = \cdots = \mathbf{V}[T_2] = n^2 \cdot \mathbf{V}[M_n] = n^2 \cdot \frac{1}{(\lambda n)^2} = \frac{1}{\lambda^2}$$

$\Rightarrow$ T_1 is a better estimator than T_2 (for $n > 1$)

Reminder: Secretary Problem



Only the local maxima (green bars) are relevant
$\Rightarrow$ Their frequency is slowing down over time
(**Exercise:** What is the expected number of local maxima?)

Example 3

Consider the secretary problem, where the ranking of the n candidates is a random permutation. What is the expected number of "best-so-far" candidates?

Answer

- Let I_k be an indicator random variable which is one iff the k-th secretary is "best-so-far"
- Side Remark: It turns out that the set of random variables $I_1, I_2, \ldots, I_n$ are independent, but this requires a proof (see Exercise Sheet) and we won't need it here!
- We have $\mathbf{P}[\, I_k = 1\,] = 1/k$, as the ranking of the first k secretaries is a random permutation over k elements
- Hence with $I := \sum_{k=1}^{n} I_k$, we have

$$\begin{aligned}\mathbf{E}[\, I\,] = \sum_{k=1}^{n} \mathbf{E}[\, I_k\,] &= \sum_{k=1}^{n} \mathbf{P}[\, I_k = 1\,] \\ &= \sum_{k=1}^{n} 1/k \approx \log(n).\end{aligned}$$

- This solves the question, but in relation to the optimal algorithm presented in Lec. 12, we can also see from the above derivation that:

$$\sum_{k=n/e+1}^{n} \mathbf{E}[\, I_k\,] = \sum_{k=n/e+1}^{n} \frac{1}{k} = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=n/e}^{n} \frac{1}{k} \approx \log(n) - \log(n/e) = \log(e) = 1,$$

$\Rightarrow$ expected number of "best-so-far" candidates among $\{n/e+1, \ldots, n\}$ is exactly one.

Example 3

Consider the secretary problem, where the ranking of the n candidates is a random permutation. What is the expected number of "best-so-far" candidates?

Extension: What happens if the n candidates arrive according to a "worst-case" permutation?

- Let I_k be an indicator random variable which is one iff the k-th secretary is "best-so-far"
- Side Remark: It turns out that the set of random variables $I_1, I_2, \ldots, I_n$ are independent, but this requires a proof (see Exercise Sheet) and we won't need it here!
- We have $\mathbf{P}[\, I_k = 1 \,] = 1/k$, as the ranking of the first k secretaries is a random permutation over k elements
- Hence with $I := \sum_{k=1}^{n} I_k$, we have

$$\mathbf{E}[\, I \,] = \sum_{k=1}^{n} \mathbf{E}[\, I_k \,] = \sum_{k=1}^{n} \mathbf{P}[\, I_k = 1 \,]$$
$$= \sum_{k=1}^{n} 1/k \approx \log(n).$$

- This solves the question, but in relation to the optimal algorithm presented in Lec. 12, we can also see from the above derivation that:

$$\sum_{k=n/e+1}^{n} \mathbf{E}[\, I_k \,] = \sum_{k=n/e+1}^{n} \frac{1}{k} = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=n/e}^{n} \frac{1}{k} \approx \log(n) - \log(n/e) = \log(e) = 1,$$

$\Rightarrow$ expected number of "best-so-far" candidates among $\{n/e + 1, \ldots, n\}$ is exactly one.