INSE6300 Jia Yuan Yu

1: Data

Concordia February 9, 2016

This course is about the interplay between performance metrics, decisions, and guarantees. We consider two cases: with and without uncertainty.

1 Without uncertainty

First, determine the possible values A of the decision variable a: technology A, B, engine size X, Y. Next, fix a performance metric $Q : A \to \mathbb{R}$, e.g., pollution output of the engine. If Q is unknown, measure $Q(a)$ for every possible value of $a \in A$.

We can then easily answer questions of the form: Given a threshold λ, what values of a guarantee that $Q(a) \geq \lambda$?

Observe that this approach requires $|A|$ measurements.

2 With uncertainty

In many supply chain situations, however, we do not have a deterministic performance metric Q. We produce a sequence of identical items: item 1, item 2, etc. We then measure the quality of each item, and denote these measurements (or observations or data) by $X_1, X_2, \ldots$ Even if there is only a single possible decision, these observations display a certain uncertainty.

Consider, for instance, the following examples:

- customer satisfaction in support center,
- quality of diamonds in production line,
- weight of apples on a farm.

How do we give performance guarantees in the presence of uncertainty?

3 Review of Statistics

Probability theory and statistics is the most widely used set of mathematical tools for modeling uncertainty.

- Probability space,
  - Set of outcomes Ω,
  - Set of events $\mathcal{F}$ containing subsets of Ω,

  - A function (probability measure) $P : \mathcal{F} \to \mathbb{R}$.

- Random variables, distribution functions,
  - Real-value random variable is a (measurable) function $X : \Omega \to \mathbb{R}$.
  - Probability distribution function or cumulative distribution function: $F(x) = P(X \leq x)$ for all x.
  - Probability density function, if it exists: $F(x) = \int_{-\infty}^{x} f(z)dz$.
- Examples (Bernoulli, Normal, Uniform, Exponential etc.),
  - Bernoulli with parameter p: $P(X = 1) = p$, $P(X = 0) = 1 - p$.
  - Exponential with rate λ: $F(x) = (1 - e^{-x})1_{[x \geq 0]}$ for all $x \in (-\infty, \infty)$.
  - Normal $N(0, 1)$: $f(x) = (2\pi)^{-1/2} e^{-x^2/2}$ for all $x \in (-\infty, \infty)$.
- Expectation,
  - Bernoulli: $\mathbb{E}X = p * 1 + (1 - p) * 0$

Statistics is the study of data using probability theory: we have access to random variables $X_1, X_2, \ldots$, but we don't know their distributions. We want to use $X_1, X_2, \ldots$ to infer their distributions.

- Classical: Let P denote the joint distribution of $X_1, X_2, \ldots, X_n$. We assume that P belongs to a known set $\{P_\theta : \theta \in \Theta\}$. The goal is to find the true value θ^* or a subset containing it.
- Bayesian: Assume that θ^* is a random variable from a known distribution.

Example 3.1. Consider $X_1, X_2, \ldots, X_n$ corresponding to the measured lifespans of n light bulbs (no decisions involved). Assume that they are independent and identically distributed according to a normal distribution P. Estimate the mean of P.

3.1 Probabilistic guarantees

Suppose that for a given decision a, we observe the following sequence of performance measurements: $X_1^a, X_2^a, \ldots, X_n^a$. Suppose that these measurements are i.i.d. with distribution F. Given δ, we would like to find an λ give guarantees of the form:

$$\mathbb{P}(X_{n+1}^a > \lambda) \geq 1 - \delta. \tag{1}$$

Observe that the above guarantee is equivalent to $1 - F(\lambda) \geq 1 - \delta$ or $\lambda \leq F^{-1}(\delta)$.

One approach is to find an estimate $\hat{F}_n$ of the distribution F using the data $X_1^a, X_2^a, \ldots, X_n^a$. If $\hat{F}$ is a very good estimate of F, then we can say that

$$\mathbb{P}(X_{n+1}^a > \hat{F}_n^{-1}(\delta)) \geq 1 - \delta.$$

Remark 1 (Relation to Newsboy problem). Recall that F^{-1} appears also in the solution to the Newsboy problem. There, the distribution F is assumed to be known, whereas in this course, we need to estimate F from data.

3.2 Estimating distributions

Figure 1: From `http://candywow.weebly.com/`

Estimation with candies. Suppose that your supply chain produces candies and that the color of each candy corresponds to a quality value:

- Green = 1
- Yellow = 2
- Orange = 3
- Red = 3
- Purple = 5

Students observe $X_1, X_2, \ldots$ The (unknown) true distribution F is

- $\mathbb{P}(X_i \leq 1) = 21/95$
- $\mathbb{P}(X_i \leq 2) = (21 + 15)/95$
- $\mathbb{P}(X_i \leq 3) = (21 + 15 + 49)/95$
- $\mathbb{P}(X_i \leq 5) = (21 + 15 + 49 + 10)/95$

We can estimate the distribution F for a sequence of i.i.d. random variables as follows. Let $X_1, X_2, \ldots, X_n$ denote the samples. Construct the following empirical distribution function, for every x:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{[X_i \leq x]}.$$

How well does $\hat{F}_n$ estimate F? Good news: exceptionally well!

Theorem 3.1 (Dvoretzky–Kiefer–Wolfowitz Inequality[1])**.** *For every $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

Remark 2. The ε above is analogous to the notion of "margin of error," whereas $1-\delta$ is analogous to "confidence."

Homework: Combine DKW Inequality with (2).

3.3 Estimating normal distributions

Normal random variables are entirely characterized by the mean μ and variance σ^2. The following sample-mean is an unbiased mean estimator:

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

The following is an unbiased variance estimator:

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \hat{\mu}_n)^2.$$

How well do $\hat{\mu}_n$ and $\hat{\sigma}_n^2$ estimate μ and σ^2? Very well, thank you!

Theorem 3.2 (Hoeffding)**.** *Let $X_1, X_2, \ldots$ be i.i.d. random variables that take values in the interval $[a, b]$, and have mean μ. Let*

$$\hat{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then, for every n and $\varepsilon > 0$:

$$\mathbb{P}\left(\left|\hat{X}_n - \mu\right| \geq \varepsilon\right) \leq 2\exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right).$$

3.4 Multiple decisions

When the decision maker is faced with a set A of possible decisions, we have a different sequence of measurements for each decision $a \in A$:

$$X_1^a, X_2^a, \ldots$$

Given $\delta > 0$, we can find a set of values $\{\lambda^a \mid a \in A\}$ such that

$$\mathbb{P}(X_{n+1}^a > \lambda^a) \geq 1 - \delta. \tag{2}$$

We can then answer questions of the form: What decisions $a \in A$ guarantee that an arbitrary lightbulb has a lifespan above γ with probability 0.99?

Remark 3. Observe that for $\gamma < \lambda^a$, we have $\{X > \lambda^a\} \subseteq \{X > \gamma\}$, and hence $\mathbb{P}(X > \gamma) \geq \mathbb{P}(X > \lambda^a)$.

[1] Dvoretzky, A.; Kiefer, J.; Wolfowitz, J. (1956), "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator", Annals of Mathematical Statistics 27 (3): 642–669.

4 Control charts

Next week, we look at the following problem. Suppose that you produce integrated circuits (computer chips), voltage measurements can be taken on these chips. Many things can go wrong in the supply chain (silicon wafer, photoresist, etching, etc.). How do we quickly detect that something went wrong somewhere?

Suppose that the measurements are $X_1, X_2, \ldots, X_n$ and i.i.d. We assume that the mean μ and the variance σ^2 are known. The Shewhart $\bar{X}$ chart simply raises an alarm when the empirical average $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ falls outside the region

$$\left[\mu - \frac{3\sigma}{\sqrt{n}}, \mu + \frac{3\sigma}{\sqrt{n}}\right]$$

of three times the standard deviation $\sigma/\sqrt{n}$ around the mean μ.

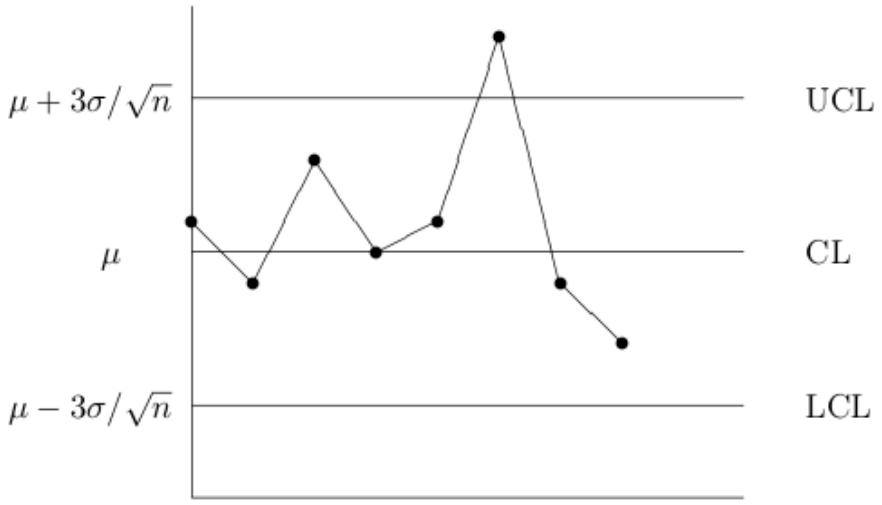

Figure 4.1: Shewhart $\overline{X}$-chart with control lines.

Figure 2: From A. Di Bucchianico, Applied Statistics, Technische Universiteit Eindhoven, 2008.

How does this alarm perform? Let's look at the probability of false alarm.

4.1 Probability of false alarm

The probability of false alarm is

$$1 - \mathbb{P}(\mu - \frac{3\sigma}{\sqrt{n}} \leq \bar{X}_n \leq \mu + \frac{3\sigma}{\sqrt{n}}) = \Phi(-3) + (1 - \Phi(2)) = 0.0027.$$

5 References

- Chapter 1 of TPE.
- A. Di Bucchianico, Applied Statistics, Technische Universiteit Eindhoven, 2008.