18.650 – Fundamentals of Statistics

2. Foundations of Inference

Goals

In this unit, we introduce a mathematical formalization of statistical modeling to make a principled sense of the **Trinity of statistical inference**.

We will make sense of the following statements:

1. Estimation:

   "$\hat{p} = \bar{R}_n$ is an estimator for the proportion p of couples that turn their head to the right"

   (side question: is 64.5% also an estimator for p?)

2. Confidence intervals:

   "$[0.56, 0.73]$ is a 95% confidence interval for p"

3. Hypothesis testing:

   "We find statistical evidence that more couples turn their head to the right when kissing"

The rationale behind statistical modeling

- Let $X_1, \ldots, X_n$ be n independent copies of X.
- The goal of statistics is to learn the distribution of X.
- If $X \in \{0, 1\}$, easy! It's $\mathsf{Ber}(p)$ and we only have to learn the parameter p of the Bernoulli distribution.
- Can be more complicated. For example, here is a (partial) dataset with number of siblings (including self) that were collected from college students a few years back: 2, 3, 2, 4, 1, 3, 1, 1, 1, 1, 1, 2, 2, 3, 2, 2, 2, 3, 2, 1, 3, 1, 2, 3, . . .
- We could make no assumption and try to learn the pmf:

x	1	2	3	4	5	6	≥ 7
$\mathbb{P}(X = x)$	p_1	p_2	p_3	p_4	p_5	p_6	$\sum_{i \geq 7} p_i$

  That's 7 parameters to learn.
- Or we could assume that $X - 1 \sim \mathsf{Poiss}(\lambda)$. That's 1 parameter to learn!

Statistical model

Formal definition

Let the observed outcome of a statistical experiment be a sample $X_1, \dots, X_n$ of n i.i.d. random variables in some measurable space E (usually $E \subseteq \mathbb{R}$) and denote by $\mathbb{P}$ their common distribution. A statistical model associated to that statistical experiment is a pair

$$(E, (\mathbb{P}_\theta)_{\theta \in \Theta}),$$

where:

- E is called sample space
- $(\mathbb{P}_\theta)_{\theta \in \Theta}$ is a family of probability measures on E;
- Θ is any set, called *parameter set*.

Parametric, nonparametric and semiparametric models

- Usually, we will assume that the statistical model is *well specified*, i.e., defined such that $\mathbb{P} = \mathbb{P}_\theta$, for some $\theta \in \Theta$.
- This particular θ is called the true parameter , and is unknown: The aim of the statistical experiment is to *estimate* θ, or check it's properties when they have a special meaning $\theta > 2?$, $\theta \neq 1/2?$, $\ldots$)
- We often assume that $\Theta \subseteq \mathbb{R}^d$ for some $d \geq 1$: The model is called *parametric*.
- Sometimes we could have Θ be infinite dimensional in which case the model is called *nonparametric*.
- If $\Theta = \Theta_1 \times \Theta_2$, where Θ_1 is finite dimensional and Θ_2 is infinite dimensional: *semiparametric* model. In these models we only care to estimate the finite dimensional parameter and the infinite dimensional one is called *nuisance* parameter. We will not cover such models in this class.

Examples of parametric models

1. For n Bernoulli trials:
$$\left(\{0,1\}, (\mathsf{Ber}(p))_{p\in(0,1)}\right).$$

2. If $X_1,\ldots,X_n \overset{iid}{\sim} \mathsf{Poiss}(\lambda)$ for some unknown $\lambda > 0$,
$$\left(\mathbb{N}, (\mathsf{Poiss}(\lambda))_{\lambda>0}\right).$$

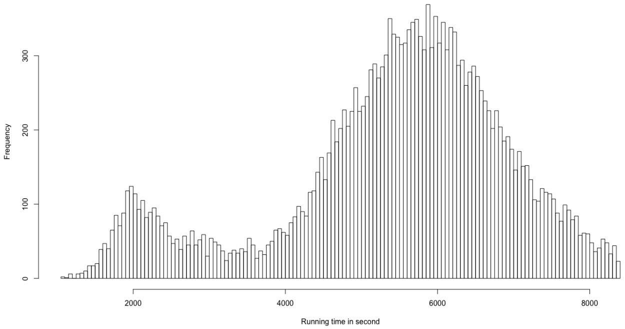
3. If $X_1,\ldots,X_n \overset{iid}{\sim} \mathcal{N}(\mu,\sigma^2)$, for some unknown $\mu \in \mathbb{R}$ and $\sigma^2 > 0$:
$$\left(\mathbb{R}, \left(\mathcal{N}(\mu,\sigma^2)\right)_{(\mu,\sigma^2)\in\mathbb{R}\times(0,\infty)}\right).$$

4. If $X_1,\ldots,X_n \overset{iid}{\sim} \mathcal{N}_d(\mu, I_d)$, for some unknown $\mu \in \mathbb{R}^d$:
$$\left(\mathbb{R}^d, (\mathcal{N}_d(\mu, I_d))_{(\mu\in\mathbb{R})}\right).$$

Mixture of Gaussians

We now introduce a more sophisticated model: Mixtures of Gaussians.

Consider the following histogram of of running times (in seconds) for the 2017 Cherry Blossom run in D.C:



There are two races: 10 mile and 5k, each corresponding to a sub-population.

Sub-populations

Assume that each sub-population is Gaussian:

$$\mathcal{N}(\mu_1, \sigma_1^2) \qquad \text{and} \qquad \mathcal{N}(\mu_2, \sigma_2^2)$$

We also need to specify the size (or proportion) of each sub-population:

- $\pi \in (0, 1)$: proportion of first sub-population
- $1 - \pi$: proportion of second sub-population

Probability density function

Pdf of a mixture of two Gaussians:

$$f(x) = \pi \cdot \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(- \frac{(x-\mu_1)^2}{2\sigma_1^2} \right) + (1-\pi) \cdot \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(- \frac{(x-\mu_2)^2}{2\sigma_2^2} \right)$$

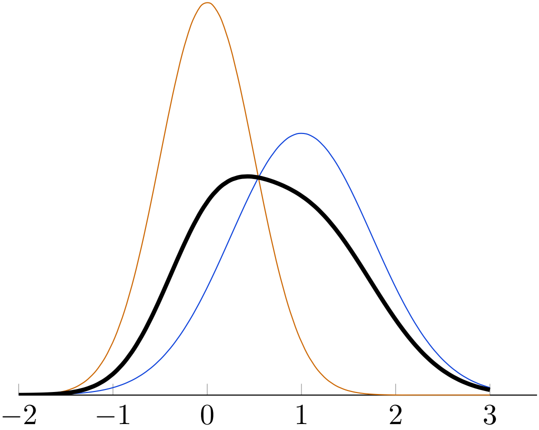


Figure 1: Mixture of $\mathcal{N}(0, 0.5^2)$ and $\mathcal{N}(1, 0.75^2)$ with $\pi = .3$

Probability density function

Pdf of a mixture of two Gaussians:

$$f(x) = \pi \cdot \frac{1}{\sigma_1\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\Big) + (1-\pi)\cdot \frac{1}{\sigma_2\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\Big)$$

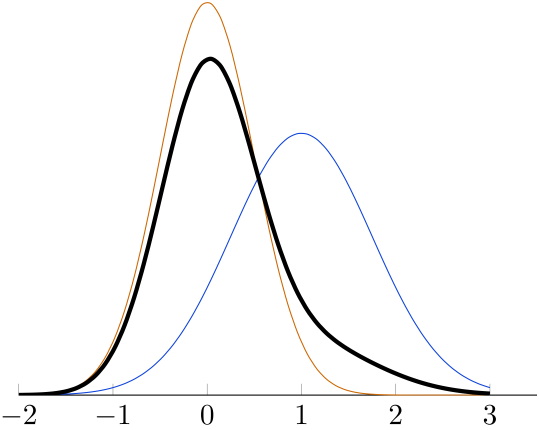


Figure 2: Mixture of $\mathcal{N}(0, 0.5^2)$ and $\mathcal{N}(1, 0.75^2)$ with $\pi = .8$

Probability density function

Pdf of a mixture of two Gaussians:

$$f(x) = \pi \cdot \frac{1}{\sigma_1\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\Big) + (1-\pi) \cdot \frac{1}{\sigma_2\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\Big)$$

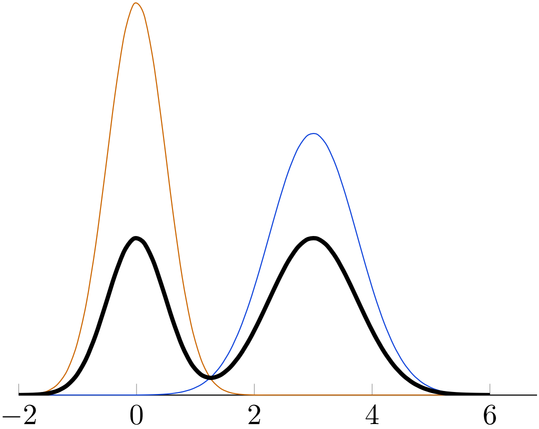
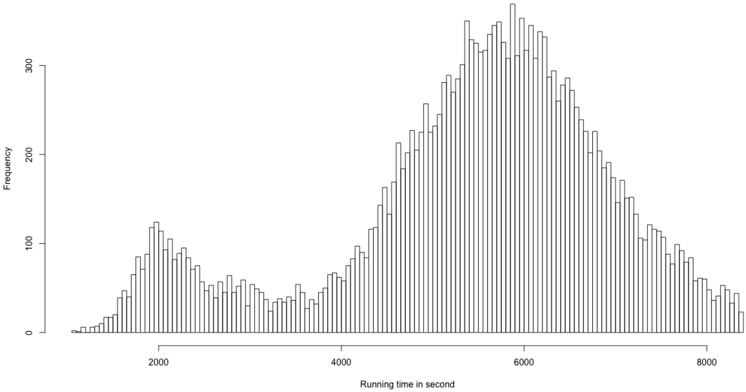
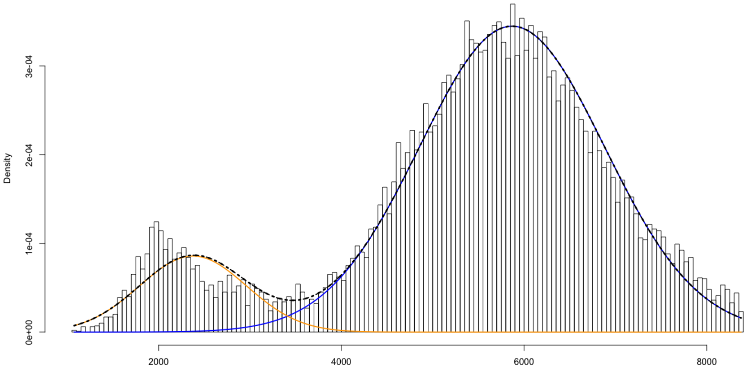


Figure 3: Mixture of $\mathcal{N}(0, 0.5^2)$ and $\mathcal{N}(3, 0.75^2)$ with $\pi = .4$

Cherry blossom



Cherry blossom



Sampling from a mixture of Gaussians

A simple way to understand mixture of Gaussians is to proceed in two steps:

1. Sample the *latent* variable $Z \sim \text{Ber}(\pi)$
2. Sample $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ independent of Z.
3. Define
$$X = ZX_1 + (1 - Z)X_2$$

One can check that X has the correct pdf.

Note that we can sample only one X_1 (if $Z = 1$) or X_2 (if $Z = 0$).

Mixture of Gaussian model

We may consider many scenarios for a model with mixtures of Gaussians, for example:

1. All five parameters unknown:

$$\left(\mathbb{R}, \left(\pi \cdot \mathcal{N}(\mu_1, \sigma_1^2) + (1-\pi) \cdot \mathcal{N}(\mu_2, \sigma_2^2)\right)_{\substack{\pi \in (0,1) \\ \mu_1, \mu_2 \in \mathbb{R} \\ \sigma_1^2, \sigma_2^2 \in (0,\infty)}}\right).$$

2. Known variances (say $\sigma_1^2, \sigma_2^2 = 1$):

$$\left(\mathbb{R}, \left(\pi \cdot \mathcal{N}(\mu_1, 1) + (1-\pi) \cdot \mathcal{N}(\mu_2, 1)\right)_{\substack{\pi \in (0,1) \\ \mu_1, \mu_2 \in \mathbb{R}}}\right).$$

3. Only unknown means (say $\sigma_1^2, \sigma_2^2 = 1$ and $\pi = \frac{1}{2}$):

$$\left(\mathbb{R}, \left(.5 \cdot \mathcal{N}(\mu_1, 1) + .5 \cdot \mathcal{N}(\mu_2, 1)\right)_{\mu_1, \mu_2 \in \mathbb{R}}\right).$$

Examples of nonparametric models

1. If $X_1, \ldots, X_n \in \mathbb{R}$ are i.i.d with unknown *unimodal*[1] pdf f:

$$\begin{aligned} E &= \mathbb{R} \\ \Theta &= \{\text{uniumodal pdf } f\} \\ \mathbb{P}_\theta = \mathbb{P}_f &= \text{distribution with pdf} f \end{aligned}$$

2. If $X_1, \ldots, X_n \in [0, 1]$ are i.i.d with unknown invertible cdf F.

$$\left([0,1], (\mathbb{P}_F)_{\text{cdf } F \text{ such that } F^{-1} \text{ exists}}\right).$$

[1] Increases on $(-\infty, a)$ and then decreases on (a, ∞) for some $a > 0$.

Identifiability

The parameter θ is called *identifiable* iff the map $\theta \in \Theta \mapsto \mathbb{P}_\theta$ is injective, i.e.,

$$\theta \neq \theta' \Rightarrow \mathbb{P}_\theta \neq \mathbb{P}_{\theta'}$$

or equivalently:

$$\mathbb{P}_\theta = \mathbb{P}_{\theta'} \Rightarrow \theta = \theta'$$

Examples

1. In all four previous examples, the parameter is identifiable.
2. If $X_i = \mathbb{1}_{Y_i \geq 0}$ (indicator function), $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, for some unknown $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, are unobserved: μ and σ^2 are not identifiable (but $\theta = \mu/\sigma$ is).

Estimation

Parameter estimation

- *Statistic*: Any measurable[2] function of the sample, e.g., $\bar{X}_n, \max_i X_i$, $X_1 + \log(1 + |X_n|)$, sample variance, etc...
- *Estimator* of θ: Any statistic whose expression does not depend on θ.
- An estimator $\hat{\theta}_n$ of θ is *weakly* (resp. *strongly*) if

$$\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P} \text{ (resp. } a.s.)} \theta \quad \text{(w.r.t. } \mathbb{P}_\theta).$$

- An estimator $\hat{\theta}_n$ of θ is *asymptotically normal* if

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, \sigma^2)$$

  The quantity σ^2 is then called *asymptotic* variance

[2] Rule of thumb: if you can compute it exactly once given data, it is measurable. You may have some issues with things that are implicitly defined such as $\sup$ or $\inf$ but not in this class

Bias of an estimator

- *Bias* of an estimator $\hat{\theta}_n$ of θ:

$$\text{bias}(\hat{\theta}_n) = \mathbb{E}\left[\hat{\theta}_n\right] - \theta.$$

- If $\text{bias}(\hat{\theta}) = 0$, we say that $\hat{\theta}$ is unbiased
- Example: Assume that $X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p)$ and consider the following estimators for p:
  - $\hat{p}_n = \bar{X}_n$: $\text{bias}(\hat{p}_n) = 0$
  - $\hat{p}_n = X_1$: $\text{bias}(\hat{p}_n) = 0$
  - $\hat{p}_n = \dfrac{X_1 + X_2}{2}$: $\text{bias}(\hat{p}_n) = 0$
  - $\hat{p}_n = \sqrt{\mathbb{1}(X_1 = 1, X_2 = 1)} \sim \text{Ber}(p^2)$: $\text{bias}(\hat{p}_n) = p^2 - p$

Variance of an estimator

An estimator is a random variable so we can compute its variance. We recall the shortcut fomula:

$$\mathsf{var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

In the previous examples:

- $\hat{p}_n = \bar{X}_n$: $\mathsf{var}(\hat{p}_n) = \frac{p(1-p)}{n}$
- $\hat{p}_n = X_1$: $\mathsf{var}(\hat{p}_n) = p(1-p)$
- $\hat{p}_n = \dfrac{X_1 + X_2}{2}$: $\mathsf{var}(\hat{p}_n) = \frac{p(1-p)}{2}$
- $\hat{p}_n = \sqrt{\mathbb{I}(X_1 = 1, X_2 = 2)} \sim \mathsf{Ber}(p^2)$: $\mathsf{var}(\hat{p}_n) = p^2(1-p^2)$

Quadratic risk

- We want estimators to have low bias and low variance at the same time.

- The *Risk* (or *quadratic risk*) of an estimator $\hat{\theta}_n \in \mathbb{R}$ is

$$
\begin{aligned}
R(\hat{\theta}_n) &= \mathbb{E}\left[|\hat{\theta}_n - \theta|^2\right] \\
&= \mathbb{E}\left[|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n] + \mathbb{E}[\hat{\theta}_n] - \theta|^2\right] \\
&= \mathbb{E}\left[|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]|^2\right] + \mathbb{E}\left[|\mathbb{E}[\hat{\theta}_n] - \theta|^2\right] + 2\mathbb{E}\left[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])(\mathbb{E}[\hat{\theta}_n] - \theta)\right] \\
&= \mathsf{var}(\hat{\theta}_n) + \mathsf{bias}^2(\hat{\theta}_n) + 2 * 0
\end{aligned}
$$

- Low quadratic risk means that both bias and variance are small:

$$\text{quadratic risk} = \text{bias}^2 + \text{variance}$$

Exercises

Let $X_1, X_2 \ldots, X_n$ be a random sample from $\mathcal{U}([a, a+1])$.

a) Find $\mathbb{E}\left[\bar{X}_n\right]$
(answer: $a + \frac{1}{2}$)
b) Is $\bar{X}_n - \frac{1}{2}$ an unbiased estimator for a?
(answer:Yes)
c) Find the variance of $\bar{X}_n - \frac{1}{2}$.
(answer: $\text{var}(\bar{X}_n - \frac{1}{2}) = \text{var}(\bar{X}_n) = \frac{\text{var}(X_1)}{n} = \frac{1}{12n}$)
d) Find the quadratic risk of $\bar{X}_n - \frac{1}{2}$. (answer: bias=0 so the answer is the same as in c)).

Confidence intervals

Confidence intervals

Let $(E, (\mathbb{P}_\theta)_{\theta \in \Theta})$ be a statistical model based on observations $X_1, \ldots, X_n$, and assume $\Theta \subseteq \mathbb{R}$. Let $\alpha \in (0,1)$.

- *Confidence interval (C.I.) of level* $1-\alpha$ for θ: Any random (depending on $X_1, \ldots, X_n$) interval $\mathcal{I}$ whose boundaries do not depend on θ and such that[3]:

$$\mathbb{P}_\theta\left[\mathcal{I} \ni \theta\right] \geq 1-\alpha, \qquad \forall \theta \in \Theta.$$

- *C.I. of asymptotic level* $1-\alpha$ for θ: Any random interval $\mathcal{I}$ whose boundaries do not depend on θ and such that:

$$\lim_{n\to\infty} \mathbb{P}_\theta\left[\mathcal{I} \ni \theta\right] \geq 1-\alpha, \qquad \forall \theta \in \Theta.$$

[3] $\mathcal{I} \ni \theta$ means that $\mathcal{I}$ contains θ. This notation emphasizes the randomness of $\mathcal{I}$ but we can equivalently write $\theta \in \mathcal{I}$.

A confidence interval for the kiss example

- Recall that we observe $R_1, \ldots, R_n \overset{iid}{\sim} \mathsf{Ber}(p)$ for some unknown $p \in (0,1)$.
- Statistical model: $\left(\{0,1\}, (\mathsf{Ber}(p))_{p\in(0,1)}\right)$.
- Recall that our estimator for p is $\hat{p} = \bar{R}_n$.
- From CLT:

$$\sqrt{n}\frac{\bar{R}_n - p}{\sqrt{p(1-p)}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)$$

  This means (precisely) that:

  - $\Phi(x)$: cdf of $\mathcal{N}(0,1)$; $\Phi_n(x)$: cdf of $\sqrt{n}\frac{\bar{R}_n - p}{\sqrt{p(1-p)}}$.
  - Then: $\Phi_n(x) \approx \Phi(x)$ (CLT) when n becomes large. Hence, for all $x > 0$,

$$\mathbb{P}\left[|\bar{R}_n - p| \geq x\right] \simeq 2\left(1 - \Phi\left(\frac{x\sqrt{n}}{\sqrt{p(1-p)}}\right)\right).$$

Confidence interval?

- For a fixed $\alpha \in (0,1)$, if $q_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of $\mathcal{N}(0,1)$, then with probability $\simeq 1-\alpha$ (if n is large enough !),

$$\bar{R}_n \in \left[p - \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}, p + \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

- More precisely

$$\lim_{n\to\infty} \mathbb{P}\left(\left[\bar{R}_n - \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}, \bar{R}_n + \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right] \ni p\right) = 1-\alpha$$

- But this is **not** a confidence interval because it depends on p!
- To fix this, there are 3 solutions.

Solution 1: Conservative bound

- Note that no matter the (unknown) value of p,

$$p(1-p) \leq 1/4$$

- Hence, asymptotically with probability at least $1-\alpha$,

$$\bar{R}_n \in \left[p - \frac{q_{\alpha/2}}{2\sqrt{n}}, p + \frac{q_{\alpha/2}}{2\sqrt{n}}\right].$$

- We get the asymptotic confidence interval:

$$\mathcal{I}_{\textsf{conserv}} = \left[\bar{R}_n - \frac{q_{\alpha/2}}{2\sqrt{n}}, \bar{R}_n + \frac{q_{\alpha/2}}{2\sqrt{n}}\right]$$

- Indeed

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{I}_{\textsf{conserv}} \ni p) \geq 1-\alpha$$

Solution 2: Solving the (quadratic) equation for p

- We have the system of two inequalities in p:

$$\bar{R}_n - \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}} \le p \le \bar{R}_n + \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}$$

- Each is a quadratic inequality in p of the form

$$(p - \bar{R}_n)^2 \le \frac{q_{\alpha/2}^2 p(1-p)}{n}$$

  We need to find the roots $p_1 < p_2$ of

$$\big(1 + \frac{q_{\alpha/2}^2}{n}\big)p^2 - \big(2\bar{R}_n + \frac{q_{\alpha/2}^2}{n}\big)p + \bar{R}_n^2 = 0$$

- This leads to a new confidence interval $\mathcal{I}_{\mathsf{solve}} = [p_1, p_2]$ such that:

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{I}_{\mathsf{solve}} \ni p) = 1 - \alpha$$

  (it's complicated to write in generic way so let us wait to have values for n, α and $\bar{R}_n$ to plug-in)

Solution 3: plug-in

- Recall that by the LLN $\hat{p} = \bar{R}_n \xrightarrow[n\to\infty]{\mathbb{P},\text{a.s.}} p$
- So by Slutsky, we also have

$$\sqrt{n}\frac{\bar{R}_n - p}{\sqrt{p(1-p)}} = \sqrt{n}\frac{\bar{R}_n - p}{\sqrt{\hat{p}(1-\hat{p})}} \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{p(1-p)}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)$$

- This leads to a new confidence interval:

$$\mathcal{I}_{\text{plug-in}} = \left[\bar{R}_n - \frac{q_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \bar{R}_n + \frac{q_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

such that

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{I}_{\text{plug-in}} \ni p) = 1-\alpha$$

95% asymptotic CI for the kiss example

Recall that in the kiss example we had $n = 124$ and $\bar{R}_n = 0.645$.
Assume $\alpha = 5\%$.
For $\mathcal{I}_{\text{solve}}$, we have to find the roots of:

$$1.03p^2 - 1.32p + 0.41 = 0 \qquad p_1 = 0.558, p_2 = 0.724$$

We get the following confidence intervals of asymptotic level 95%:

- $\mathcal{I}_{\text{conserv}} = \left[0.56\,,\,0.73\right]$
- $\mathcal{I}_{\text{solve}} = \left[0.56\,,\,0.72\right]$
- $\mathcal{I}_{\text{plug-in}} = \left[0.56\,,\,0.73\right]$

There are many[4] other possibilities in softwares even ones that use the exact distribution of $n\bar{R}_n \sim \mathsf{Bin}(n, p)$

$$\mathcal{I}_{\text{R default}} = \left[0.55\,,\,0.73\right]$$

[4] See R. Newcombe (1998). *Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods*.

Another example: The T

Photo from User: IIconic Rails on YouTube (c) Mikay Royce

Statistical problem

- You observe the times (in minutes) between arrivals of the T at Kendall: $T_1, \ldots, T_n$.
- You **assume** that these times are:
  - Mutually independent
  - Exponential random variables with common parameter $\lambda > 0$.
- You want to *estimate* the value of λ, based on the observed arrival times.

Discussion of the modeling assumptions

- Mutual independence of $T_1, \ldots, T_n$: plausible but not completely justified (often the case with independence).
- $T_1, \ldots, T_n$ are exponential r.v.: **lack of memory** of the exponential distribution:

  $$\mathbb{P}[T_1 > t + s | T_1 > t] = \mathbb{P}[T_1 > s], \quad \forall s, t \geq 0.$$

  Also, $T_i > 0$ almost surely!
- The exponential distributions of $T_1, \ldots, T_n$ have the same parameter: in average all the same inter-arrival time. True only for limited period (rush hour $\neq$ 11pm).

Estimator

- Density of T_1:
$$f(t) = \lambda e^{-\lambda t}, \quad \forall t \geq 0.$$
- $\mathbb{E}[T_1] = \dfrac{1}{\lambda}$.
- Hence, a natural estimate of $\dfrac{1}{\lambda}$ is
$$\bar{T}_n := \frac{1}{n} \sum_{i=1}^{n} T_i.$$
- A natural estimator of λ is
$$\hat{\lambda} := \frac{1}{\bar{T}_n}.$$

First properties

- By the LLN's,

$$\bar{T}_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbb{P}} \frac{1}{\lambda}$$

- Hence,

$$\hat{\lambda} \xrightarrow[n\to\infty]{\text{a.s.}/\mathbb{P}} \lambda.$$

- By the CLT,

$$\sqrt{n}\,\left(\bar{T}_n - \frac{1}{\lambda}\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, \lambda^{-2}).$$

- How does the CLT transfer to $\hat{\lambda}$? How to find an asymptotic confidence interval for λ ?

The Delta method

Let Z_n be a sequence of random variables such that

$$\sqrt{n}(Z_n - \theta) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, \sigma^2),$$

for some $\theta \in \mathbb{R}$ and $\sigma^2 > 0$ (asymptotically normal).

Let $g : \mathbb{R} \to \mathbb{R}$ be continuously differentiable at the point θ. Then,

- $g(Z_n) \xrightarrow[n\to\infty]{\mathbb{P}}$
- $(g(Z_n))_{n\geq 1}$ is also asymptotically normal;
- More precisely,

$$\sqrt{n}\,(g(Z_n) - g(\theta)) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, g'(\theta)^2\sigma^2).$$

Consequence of the Delta method

- $\sqrt{n}\ \left(\hat{\lambda} - \lambda\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, \lambda^2)$.

- Hence, for $\alpha \in (0,1)$ and when n is large enough,

$$|\hat{\lambda} - \lambda| \leq \frac{q_{\alpha/2}\lambda}{\sqrt{n}}.$$

  with probability approximately $1 - \alpha$.

- Can $\left[\hat{\lambda} - \frac{q_{\alpha/2}\lambda}{\sqrt{n}}, \hat{\lambda} + \frac{q_{\alpha/2}\lambda}{\sqrt{n}}\right]$ be used as an asymptotic confidence interval for λ ?
  No ! It depends on λ...

Three “solutions”

1. The conservative bound: we have no a priori way to bound λ
2. We can solve for λ:

$$|\hat{\lambda} - \lambda| \leq \frac{q_{\alpha/2}\lambda}{\sqrt{n}} \iff \lambda\left(1 - \frac{q_{\alpha/2}}{\sqrt{n}}\right) \leq \hat{\lambda} \leq \lambda\left(1 + \frac{q_{\alpha/2}}{\sqrt{n}}\right)$$

$$\iff \hat{\lambda}\left(1 + \frac{q_{\alpha/2}}{\sqrt{n}}\right)^{-1} \leq \lambda \leq \hat{\lambda}\left(1 - \frac{q_{\alpha/2}}{\sqrt{n}}\right)^{-1}.$$

It yields

$$\mathcal{I}_{\mathsf{solve}} = \left[\hat{\lambda}\left(1 + \frac{q_{\alpha/2}}{\sqrt{n}}\right)^{-1}, \hat{\lambda}\left(1 - \frac{q_{\alpha/2}}{\sqrt{n}}\right)^{-1}\right]$$

3. Plug-in yields

$$\mathcal{I}_{\mathsf{plug\text{-}in}} = \left[\hat{\lambda}\left(1 - \frac{q_{\alpha/2}}{\sqrt{n}}\right), \hat{\lambda}\left(1 + \frac{q_{\alpha/2}}{\sqrt{n}}\right)\right]$$

95% asymptotic CI for the Kendall T example

Assume that $n = 64$ and $\bar{T}_n = 6.23$ and $\alpha = 5\%$.

We get the following confidence intervals of asymptotic level 95%:

- $\mathcal{I}_{\text{solve}} = \big[0.13\,,\,0.21\big]$
- $\mathcal{I}_{\text{plug-in}} = \big[0.12\,,\,0.20\big]$

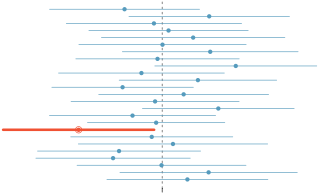
Meaning of a confidence interval

Take $\mathcal{I}_{\text{plug-in}} = \left[0.12\,,\,0.20\right]$ for example. What is the meaning of "$\mathcal{I}_{\text{plug-in}}$ is a confidence intervals of asymptotic level 95%".

Does it mean that

$$\lim_{n\to\infty} \mathbb{P}\left(\lambda \in \left[0.12\,,\,0.20\right]\right) \geq .95?$$

There is a *frequentist* interpretation[5]: If we were to repeat this experiment (collect 64 observations) then λ would be in the resulting confidence interval about of the time (image credit: `openintro.org`).



[5] The frequentist approach is often contrasted with the Bayesian approach.

Hypothesis testing

Waiting time in the ER

- The average waiting time in the Emergency Room (ER) in the US is 30 minutes according to the CDC
- Some patients claim that the new Princeton-Plainsboro hospital has a longer waiting time. Is it true?
- Here, we collect only one sample: $X_1, \ldots, X_n$ (waiting time in minutes for n random patients) with unknown expected value $\mathbb{E}[X_1] = \mu$.
- We want to know if $\mu > 30$.

This is a

© 2018 HealthTap, All rights reserved.

Statistical formulation

- Consider the two hypotheses:

$$\begin{aligned} H_0 : &\quad \mu \leq 30 \\ H_1 : &\quad \mu > 30 \end{aligned}$$

- H_0 is the *null hypothesis*, H_1 is the *alternative hypothesis*.
- We say that we *test* H_0 *against* H_1.
- We want to decide whether to *reject* H_0 (look for evidence against H_0 in the data).

Heuristic[6]:

If $\bar{X}_n > 30$+buffer

then conclude that $\mu > 30$

[6] We will be more precise in Unit 4

Recap

- A statistical model is a pair of the form $(E, (\mathbb{P}_\theta)_{\theta \in \Theta})$ where E is the sample space and $(\mathbb{P}_\theta)_{\theta \in \Theta}$ is a family of candidate probability distributions.
- A model can be well specified and identifiable.
- The trinity of statistical inference: estimation, confidence intervals and testing
- Estimator: one value whose performance can be measured by consistency, asymptotic normality, bias, variance and quadratic risk
- Confidence intervals provide "error bars" around estimators. Their size depends on the confidence level
- Hypothesis testing: we want to ask a yes/no answer about an unknown parameter.

ATTRI UTIO SU AR

Unit 2
Slide # 32
https://www.youtube.com/watch?v=VBBeRDa_gms
Citation/Attribution -- Photo from User: IIconic Rails on YouTube (c) Mikay Royce

Unit 2
Slide # 43
https://doctors.healthtap.com/hs-fs/hubfs/waiting_room.jpgt=1511990105503&width=640&name=waiting_room.jpg
Citation/Attribution -- Image © 2018 HealthTap, All rights reserved.