Statistics Feb 17, 2021

Let $X_1, X_2, \ldots, X_n$ follow the distribution F

$$\Leftrightarrow \quad \mathbb{P}(X_i \leq x) = F(x).$$

Ex: $X \sim N(\mu, \sigma^2)$

Given samples $x_1, \ldots, x_n$, what can we say about F?

$X_1, \ldots, X_n$ denote random variables, and the underlying theory will be developed on these.

$x_1, \ldots, x_n$ denotes data - actual numbers. The data are usually inserted at the very end to compute estimates or errors, etc.

A statistical model is the set of all possible forms of F:

Ex: A parametric model

$$\mathcal{F} = \left\{ f(x; \mu, \sigma^2) = F' : f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \right\}$$

where $\underbrace{\mu \in \mathbb{R}, \ \sigma^2 > 0}_{\text{parameter space.}}$

If we are only interested in the average μ, then

μ = parameter of interest

σ^2 = nuisance parameter.

Nonparametric statistical model: the model <u>cannot</u> be described using a finite number of parameters.

<u>Ex:</u> $\mathcal{F} = \{ f(x) : f \geq 0 , \int f \, dx = 1 , \underbrace{\int |f''(x)| \, dx < \infty}_{f \text{ is not } \underline{\text{too}} \text{ wiggly.}} \}$

Parameter Estimates

$\Rightarrow$ Estimate $\mu = E(X_1)$, or some other parameter,

<u>Ex:</u> μ can be written as a function of F:

$$\mu = \int x \, f(x) \, dx$$

$$= \int x \, dF(x)$$

$$= T(F)$$

any function of F is known as a "statical function"

Point Estimation

Goal: Determine a single <u>best guess</u> as to the value of a specific parameter.

<u>Notation</u> Denote by $\theta \in \mathbb{R}^k$ $\theta = \overbrace{(\theta_1, \theta_2, \theta_3, \dots, \theta_k)}^{\text{unknown.}}$ the true values of the parameter.

Denote by $\hat{\theta}$ our estimates of $\theta_1, \dots, \theta_k$.

random variable, or once we plug in data, a number/vector.

<u>Ex:</u> $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$

$\approx E(X_1)$

How do we determine if $\hat{\theta}$ is a <u>good</u> estimator?

<u>Bias</u>: $\text{bias}(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta}) - \theta$

<u>Definition</u> $\hat{\theta}$ is <u>consistent</u> if $\hat{\theta} \xrightarrow{P} \theta$ as $n \to \infty$.

$\iff \mathbb{P}(|\hat{\theta} - \theta| > \epsilon) \to 0$ as $n \to \infty$ for all $\epsilon > 0$.

The distribution of $\hat{\theta}$ is known as the <u>sampling distribution</u>.

The standard deviation of $\hat{\theta}$ is known as the <u>standard error</u>:

$$se = se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$$

Often the se will depend on θ, so it will need to be estimated - the estimated se is $\hat{se}$.

<u>Example</u> Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. (iid r.v.'s)

Let $\hat{p} = \frac{1}{n} \sum X_i$.

$\mathbb{E}(\hat{p}) = \frac{1}{n} \sum \mathbb{E}(X_i) = p$.

Standard error:
$$\begin{aligned} se &= \sqrt{Var(\hat{p})} \\ &= \sqrt{Var\left(\frac{1}{n} \sum X_i\right)} \\ &= \sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

$$\Rightarrow \quad \hat{se} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

To obtain <u>actual</u> estimates, substitute data $x_1, \ldots, x_n$ for the random variables $X_1, \ldots, X_n$.

MSE = mean squared error

$= \mathbb{E}_\theta\left((\hat\theta-\theta)^2\right)$

Thm $\text{MSE} = \text{bias}(\hat\theta)^2 + \text{Var}_\theta(\hat\theta)$.

Proof Easy, just expand each term above.

Thm If bias $\to 0$, and se $\to 0$ as $n\to\infty$, then $\hat\theta$ is consistent: $\hat\theta \xrightarrow{P} \theta$.

Def: $\hat\theta$ is <u>asymptotically normal</u> if

$$\frac{\hat\theta-\theta}{se} \rightsquigarrow N(0,1)$$

convergence in distribution

i.e. $\hat\theta \approx N(\theta, se^2)$

Confidence Interval

A $1-\alpha$ confidence interval for θ is $C = (a, b)$

functions of the data.

such that

$$\mathbb{P}_\theta(\theta \in C) \geq 1-\alpha$$ for all θ in the parameter space.

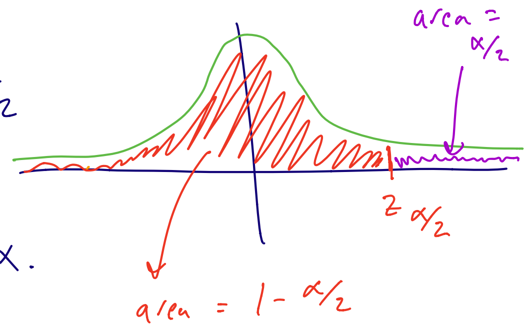
fixed random.

If our estimator $\hat\theta$ is <u>asymptotically normal</u> then we can construct a <u>normal-band</u> confidence interval.

If $\hat{\theta} \approx N(\theta, se^2)$. Let $\Phi(z) = \mathbb{P}(Z \le z)$ ← $N(0,1)$ r.v.

Set $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$

$\iff \Phi(z_{\alpha/2}) = \mathbb{P}(Z \le z_{\alpha/2}) = 1 - \alpha/2$



$\iff \mathbb{P}(-z_{\alpha/2} < Z \le z_{\alpha/2}) = 1 - \alpha.$

And since $\dfrac{\hat{\theta} - \theta}{\hat{se}} \approx N(0,1)$ then

$$\mathbb{P}\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\hat{se}} \le z_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow \mathbb{P}(\hat{\theta} - \hat{se}\, z_{\alpha/2} < \theta < \hat{\theta} + \hat{se}\, z_{\alpha/2}) \approx 1 - \alpha.$$

$$\Rightarrow C = (\hat{\theta} - \hat{se}\, z_{\alpha/2},\ \hat{\theta} + \hat{se}\, z_{\alpha/2})$$

$$\Rightarrow \mathbb{P}(\theta \in C) \to 1 - \alpha \quad \text{as } n \to \infty.$$

Empirical Distribution Function

Let $X_1, \ldots, X_n$ be IID with F as their CDF.

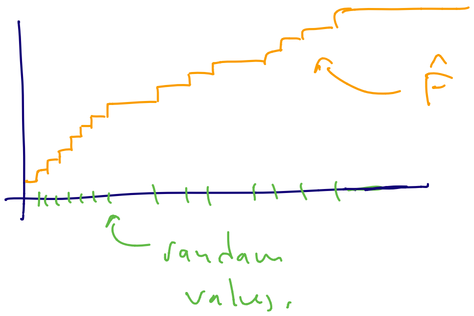
We can estimate F by:

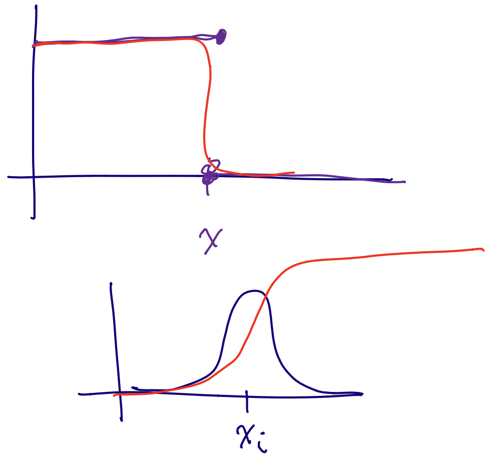
Def: The empirical distribution function is $\hat{F}$ which puts mass $\frac{1}{n}$ at each X_i:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{I(X_i \le x)}_{\substack{\text{indicator function} \\ = 1 \text{ if } X_i \le x \\ = 0 \text{ if } X_i > x}}$$

Nonparametric estimator.

Graphically





Thm: For any fixed x,

$$\mathbb{E}(\hat{F}(x)) = F(x)$$

$$\mathrm{Var}(\hat{F}(x)) = \frac{F(x)(1-F(x))}{n} \to 0$$

$$\mathrm{MSE} = \frac{F(x)(1-F(x))}{n} \to 0$$

$$\hat{F} \xrightarrow{P} F$$

Once we have constructed $\hat{F}$, it is easy to estimate any function of F.

Plugin-estimate Just evaluate $\theta = T(F)$ at $\hat{F}$

$$\Rightarrow \quad \hat{\theta} = T(\hat{F}).$$

Ex:

$$\mu = \mathbb{E}(X_i)$$

$$= \int x \, f(x) \, dx$$

$$= \int x \, dF(x)$$

$$\hat{\mu} = \int x \, d\hat{F}(x) = \frac{1}{n} \sum X_i .$$

Parametric Inference (Chapter 9 in Wasserman).

Recall that a parametric model is

$$\mathfrak{F} = \{ f(x;\theta) : \theta \in \Theta \}$$

might be a vector.

Method of Moments

The j^{th} moment of a random variable X is

$$\alpha_j = \alpha_j(\theta) = \mathbb{E}_\theta(X^j) = \int x^j f(x;\theta)\,dx$$

$$= \int x^j \, dF(x;\theta).$$

The j^{th} sample moment is just the plugin estimate of the j^{th} moment:

$$\Rightarrow \quad \hat{\alpha}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$$

If $\theta \in \mathbb{R}^k$, $\theta = (\theta_1, \dots, \theta_k)$ then the method of moments estimator $\hat{\theta}$ is the value of $\hat{\theta}$ such that

Typo in book, θ should $\hat{\theta}$.

$$\left\{ \begin{aligned} \alpha_1(\hat{\theta}) &= \hat{\alpha}_1 \\ \alpha_2(\hat{\theta}) &= \hat{\alpha}_2 \\ &\vdots \\ \alpha_k(\hat{\theta}) &= \hat{\alpha}_k . \end{aligned} \right\}$$

Equates moments with sample moments.

k equations in k unknowns.

Ex: $X_1, \dots, X_n \sim N(\mu, \sigma^2)$. We want to estimate μ and σ^2.

$\Rightarrow \mu = \mathbb{E}(X_1) = \alpha_1$

$$\sigma^2 = Var(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = \alpha_2 - \alpha_1^2.$$

$$\alpha_2 = \sigma^2 + \alpha_1^2 = \sigma^2 + \mu^2.$$

$$\hat{\alpha}_1 = \frac{1}{n} \sum X_i$$

$$\hat{\alpha}_2 = \frac{1}{n} \sum X_i^2.$$

So we must solve:

$$\hat{\mu} = \hat{\alpha}_1$$

$$\underbrace{\hat{\sigma}^2 + \hat{\mu}^2}_{\text{moments eval at } \hat{\mu}, \hat{\sigma}^2} = \underbrace{\hat{\alpha}_2}_{\text{sample moments}}$$

Solve to obtain

$$\hat{\mu} = \frac{1}{n} \sum X_i$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum X_i^2 - \left(\frac{1}{n} \sum X_i\right)^2$$

$$= \frac{1}{n} \sum (X_i - \bar{X})^2$$

$\bar{X} = \frac{1}{n} \sum X_i$

Another estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2.$$