STAT 830
Statistical Inference

Richard Lockhart

Simon Fraser University

STAT 830 — Fall 2013

What I assume you already know

- Some of this jargon

What I want you to learn

- All of this jargon.

Statistical Models pp 87-88

- **Definition**: A **model** is a family $\mathcal{F}$ of possible distributions for some random variable X. Often we will write

$$\mathcal{F} = \{P_\theta; \theta \in \Theta\}$$

or

$$\mathcal{F} = \{f(x, \theta); \theta \in \Theta\}.$$

- WARNING: Data set is X, so X will generally be a big vector or matrix or even more complicated object.
- The model is called *parametric* if Θ is finite dimensional.
- The model is called *non-parametric* if Θ is infinite dimensional.
- The model is called *semi-parametric* if a typical element θ of Θ can be thought of as a pair ϕ, ψ where ϕ is finite dimensional and ψ is infinite dimensional.

Basic Examples

- **Example**: If we specify $X_1, \ldots, X_n$ are a sample from the $N(\mu, \sigma^2)$ distribution then we are studying a (two dimensional) parametric model and X is the vector $(X_1, \ldots, X_n)$.
- **Example**: If we specify $X_1, \ldots, X_n$ are a sample from a completely unknown distribution then the parameter space is $\{F\}$, the set of *all* cdfs. This is a non-parametric model.

Example: Regression p 89

- Think about data $((X_1, Y_1), \ldots, (X_n, Y_n))$; begin by assuming the pairs are independent.
- Interest often centres on the behaviour of Y for a given value of X.
- One standard summary of this behaviour is the *regression* function

$$\mathrm{E}(Y_i|X_i)\,.$$

- There are many models — parametric, non-parametric and semi parametric — for data of this type.

Parametric Regression

- If we say that $((X_1, Y_1), \ldots, (X_n, Y_n))$ are a sample from a bivariate normal population then again we have a parametric model.
- In this model there are constants β_0 and β_1 for which

$$\mathrm{E}(Y_i|X_i) = \beta_0 + \beta_1 X_i$$

- Moreover, if we define $\epsilon_i = Y_i - \beta X_i - \beta_0$ then $\epsilon_i \perp\!\!\!\perp X_i$ and $\epsilon_i \sim N(0, \sigma^2)$ for some σ.
- There are 5 parameters: mean and variance of X, variance of ϵ and the two regression parameters.
- The model can also be parametrized using the mean and variance of X, the mean and variance of Y and the covariance (or the correlation) between X and Y.

Regression Models Continued

- Sometimes we take the X_i to be deterministic constants, $x_1, \ldots, x_n$.
- This happens in designed experiments where the X values are chosen by the experimenter.
- We still suppose
$$\mathrm{E}(Y_i|X_i) = \beta_0 + \beta_1 X_i$$
- We define ϵ_i as before and note that $\mathrm{E}(\epsilon_i) = 0$ is automatic.
- We might assume the ϵ_i are independent and identically distributed – that is, they all have the same distribution.
- If we specify nothing more the model is *semi-parametric* because the regression function model is parametric but the distributional model for the *errors* ϵ_i is non-parametric.
- Or we might consider some parametric model for the ϵ_i other than normal.

Nonparametric Regression Models

- Or we might just suppose the pairs are iid and write

$$\mathrm{E}(Y_i|X_i) \equiv \phi(X_i)$$

  where the function ϕ is some unknown continuous function. This model is *non-parametric*
- The lines between these terms are not totally clear – nor are the precise definitions important.
- To give you an idea we might suppose the ϵ_i are normally distributed but let the conditional variance, given X_i, be some unknown (or known even) function of X_i.

Model Misspecification

- Assumption in this course: true distribution P of X is P_{θ_0} for some $\theta_0 \in \Theta$.
- JARGON: θ_0 is *true value* of the parameter.
- Or in a non-parametric model F_0 is the true distribution of the data.
- Notice: in a parametric model this assumption is wrong; we hope it is not wrong in an important way.
- If it's wrong: enlarge model, put in more distributions, make Θ bigger.
- **Jargon**: The field of model *misspecification* studies the errors induced by the fact that the model is not right – does not include the true distribution of the data.

Inference goals p 89

- Observe value of X, guess θ_0 or some property of θ_0.
- Classic mathematical versions of guessing:
  1. Point estimation: compute estimate $\hat{\theta} = \hat{\theta}(X)$ which lies in Θ (or something close to Θ).
  2. Point estimation of ftn of θ: compute estimate $\hat{\phi} = \hat{\phi}(X)$ of $\phi = g(\theta)$.
  3. Interval (or set) estimation: compute set $C = C(X)$ in Θ which we think will contain θ_0.
  4. Hypothesis testing: choose between $\theta_0 \in \Theta_0$ and $\theta_0 \notin \Theta_0$ where $\Theta_0 \subset \Theta$.
  5. Prediction: guess value of an observable random variable Y whose distribution depends on θ_0. Typically Y is the value of the variable X in a repetition of the experiment.

Schools of statistical thinking

Main schools of thought summarized roughly as follows:

- **Neyman Pearson**: A statistical procedure is evaluated by its long run frequency performance. Imagine repeating the data collection exercise many times, independently. Quality of procedure measured by its average performance when true distribution of X values is P_{θ_0}.
- **Bayes**: Treat θ as random just like X. Compute conditional law of unknown quantities given knowns. In particular ask how procedure will work on the data we actually got – no averaging over data we might have got.
- **Likelihood**: Try to combine previous 2 by looking only at actual data while trying to avoid treating θ as random.

In this course we start by using Neyman Pearson approach to evaluate quality of likelihood and other methods.