STA 250/MTH 342 – Intro to Mathematical Statistics

Lecture 6

The sampling (or frequentist) view point

- Both Bayesians and frequentists agree on the need to build models for the distribution of the data given certain parameters—$f(x|\theta)$ or $p(x|\theta)$.
- Frequentists do not consider the parameter as a random quantity, but rather think of it as a *fixed unknown* quantity.
- From this perspective, the only random quantities are the data (or functions of data, i.e. statistics) *prior to the experiment.*
- Once the experiment is done and we have observed the data, then nothing is random.
- So we can no longer say things like "*Given the data*, the *chance* of ..."

Point estimation

- Let us come back to the point estimation problem—the "guessing" of the value of a parameter θ *based on observed data* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.
- Such guesses, which are *functions of the data*, are called *estimators* for the parameter. Common notations: $\hat{\theta}(\mathbf{X})$, $\delta(\mathbf{X})$, etc. Estimators are essentially "rules" that map data to guesses.
- If the observed data is $\mathbf{X} = \mathbf{x}$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, the realized value of an estimator $\delta(\mathbf{X})$ is $\delta(\mathbf{x})$, which is called an *estimate*. Estimates are the actual guesses.
- What is a criterion for "good" estimators?
- A good estimator should be such that the estimate and the actual parameter θ are *"likely to be close"*.

What does "likely to be close" mean?

2. The sampling view:
    - The parameter is a *fixed* unknown number. The only random quantities are the data.
    - After data is observed, however, nothing is random. No matter what estimator $\delta(\mathbf{X})$ we are considering, we cannot judge how close the parameter θ is to a *single realization of* the estimate $\delta(\mathbf{x})$ after $\mathbf{X} = \mathbf{x}$ is observed.
    - In this case, we want to choose an estimator $\delta(\mathbf{X})$ that will "with high probability" take values close to the underlying fixed θ.
    - Such a probabilistic statement can only be made *before the experiment*, or under repeated experiment.

Note that this is a "before the experiment" *view, in contrast to the "after the experiment" view taken by the Bayesian perspective.*

What does "likely to be close" mean?

This depends on which view about inference you are taking ...

1. The Bayesian view:
   - Both the parameter θ and data $\mathbf{X}$ are random variables.
   - *After* we have observed the data $\mathbf{X} = \mathbf{x}$, only θ is random, and its distribution is the posterior distribution $\xi(\theta|\mathbf{x})$.
   - In this case, we want to pick an estimate $\delta(\mathbf{x})$ such that *a posteriori* the parameter θ, which is random, will likely take values close to the estimate $\delta(\mathbf{x})$.

Note that here the parameter is random while the estimate $\delta(\mathbf{x})$ is a fixed number given the observed data $\mathbf{X} = \mathbf{x}$.

The sampling distribution of an estimator

- Now we take the view of a *fixed* but *unknown* parameter θ.
- The data $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are jointly distributed random variables with a joint distribution $f(\mathbf{x}|\theta)$, presumably different for each θ.
- Any estimator $\delta(\mathbf{X})$ is also a random variable, and will have a distribution $f_\delta(u|\theta)$ given θ.
- This is called the *sampling distribution* of the estimator $\delta(\mathbf{X})$.

Example: the political poll revisited

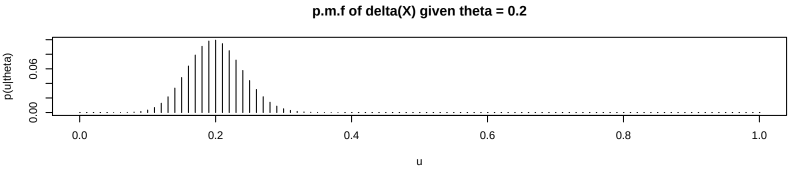
Let us consider a simple estimator,

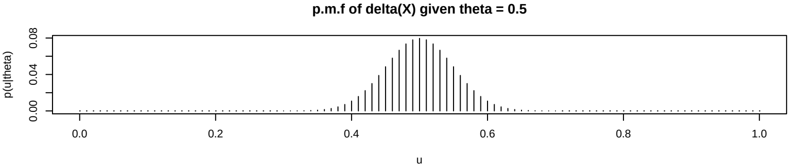
$$\delta(X) = \frac{X}{n}.$$

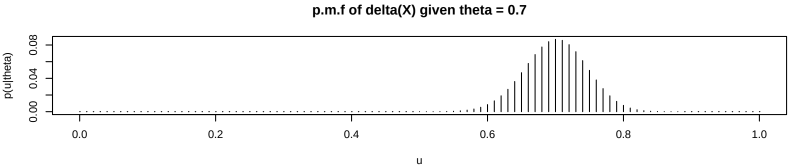
With $n = 100$, its p.m.f is

$$\begin{aligned} P\Big(\delta(X) = u \Big| \theta\Big) &= \binom{100}{100u} \theta^{100u} (1-\theta)^{100(1-u)} \quad \text{for } u = .01, .02, \ldots, 1 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Probability mass function of the estimator $\delta(X) = X/n$







Example: measuring air pollutant

If n independent readings $X_1, X_2, \ldots, X_n$ are taken from the distribution $\mathrm{N}(\theta, \tau^2)$ with known τ, then what is the sampling distribution of the *sample mean estimator* (given θ)?

$$\delta(\mathbf{X}) = \delta(X_1, X_2, \ldots, X_n) = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

Note that

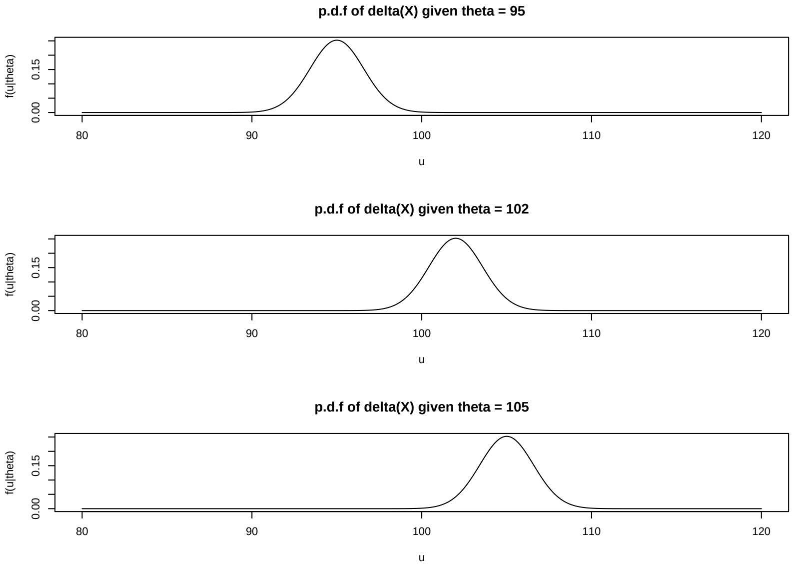
$$E[\bar{X}|\theta] = \frac{\sum_{i=1}^n E(X_i|\theta)}{n} = \frac{n\theta}{n} = \theta$$

and

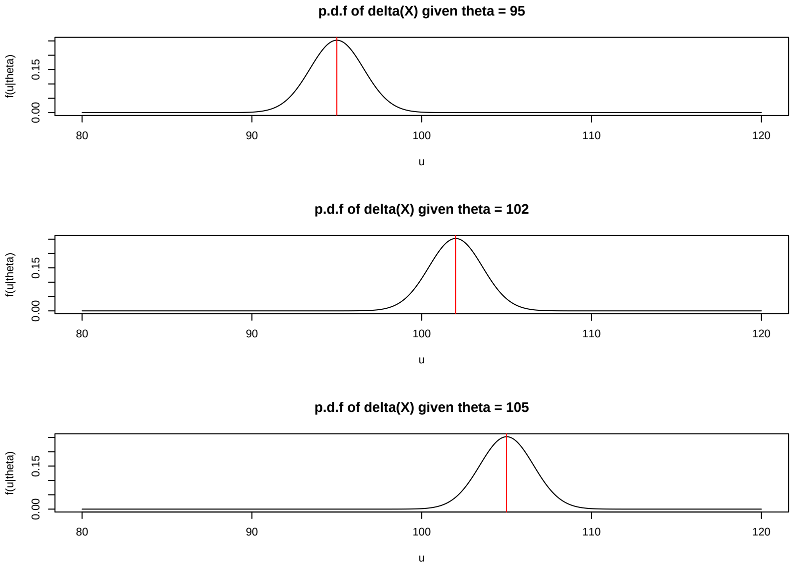
$$\mathrm{Var}(\bar{X}|\theta) = \mathrm{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\middle|\theta\right) = \frac{\sum_{i=1}^n \mathrm{Var}(X_i|\theta)}{n^2} = \frac{n\tau^2}{n^2} = \frac{\tau^2}{n}.$$

Since the sum of independent normal random variables are still normal random variables, its sampling distribution is $\mathrm{N}(\theta, \tau^2/n)$.

Sampling distribution of $\delta(\mathbf{X})$ for $n = 10$, $\tau = 5$



Sampling distribution of $\delta(\mathbf{X})$ for $n = 10$, $\tau = 5$



Criteria for selecting "good" estimators

Two desirable properties of estimators under repeated experiments:

- "Overall accuracy": $E(\delta(X)|\theta)$ is close to θ.
- "Precision": $\mathrm{Var}(\delta(X)|\theta)$ is small.

Bias of an estimator (Textbook Section 8.7)

An estimator for θ is said to be *unbiased* if its overall accurate for all possible values of θ. That is

$$E(\delta(X)|\theta) = E_\theta(\delta(X)) = E_\theta(\delta) = \theta \quad \textit{for all } \theta.$$

The *bias* of $\delta(X)$ given θ is defined to be

$$B_\delta(\theta) = E_\theta(\delta) - \theta.$$

- $B_\delta(\theta) > 0$: $\delta(X)$ tends to overestimate θ.
- $B_\delta(\theta) < 0$: $\delta(X)$ tends to underestimate θ.

Back to the political poll example

- The estimator $\delta_1(X) = \frac{X}{n}$ is unbiased.

$$E_\theta(\delta_1) = \theta \quad \text{for all } \theta.$$

- The estimator $\delta_2(X) = \frac{X+12}{n+24}$ is not unbiased.

$$B_{\delta_2}(\theta) = E_\theta(\delta_2) - \theta = \frac{n\theta + 12}{n+24} - \theta = \frac{12 - 24\theta}{n+24}.$$

So for $\theta > 1/2$, $B_{\delta_2}(\theta) < 0$ and for $\theta < 1/2$, $B_{\delta_2}(\theta) > 0$.

Question: Is $\delta_1(X)$ always more desirable than $\delta_2(X)$?

Do we always want unbiased estimators?

- Unbiasedness is a nice property.
- But unbiased estimators are not necessarily "good" estimators. (Draw a figure.)
- We need some control over the *precision* of the estimator as well.
- So it is desirable to have estimators with small variance

$$\mathrm{Var}\left(\delta(\mathbf{X})|\theta\right).$$

- Of course neither unbiasedness nor precision alone is enough. (Draw figure).
- More generally, the goal is to have estimators likely to take values close to the unknown *fixed* parameter.

More general criteria for selecting "good" estimators

Recall from last time the goal of constructing an estimator $\delta(\mathbf{X})$ so that $\delta(\mathbf{X})$ will *likely to be close to* θ.

- Again, we need a notion of distance between the estimate and the parameter.
- Can again use a loss function just like before such as
  1. The absolute error loss: $L(\theta, a) = |\theta - a|$.
  2. The squared error loss: $L(\theta, a) = (\theta - a)^2$.
  3. The step error loss: $L(\theta, a) = \mathbf{1}(|\theta - a| > \Delta)$.

  So we can again try to choose an estimator that minimizes the *expected loss*.

But now the expectation is taken over the distribution of the estimator given θ:

$$
\begin{aligned}
E(L(\theta,\delta(\mathbf{X}))|\theta) = E_\theta(L(\theta,\delta)) &= \int_{-\infty}^{\infty} L(\theta,u) f_\delta(u|\theta)\,du \\
&= \int_{-\infty}^{\infty} L(\theta,\delta(x)) f(x|\theta)\,d\theta
\end{aligned}
$$

where $f_\delta(u|\theta)$ denotes the p.d.f of the estimator $\delta(\mathbf{X})$.

This expectation is also called the *(sampling) risk function* of estimator δ:

$$R_\delta(\theta) := E_\theta(L(\theta,\delta))$$