Back to estimators...

So far, we have:

- Identified estimators for common parameters
- Discussed the sampling distributions of estimators
- Introduced ways to judge the "goodness" of an estimator (bias, MSE, etc.)
- Used estimators in confidence intervals, hypothesis testing, etc.

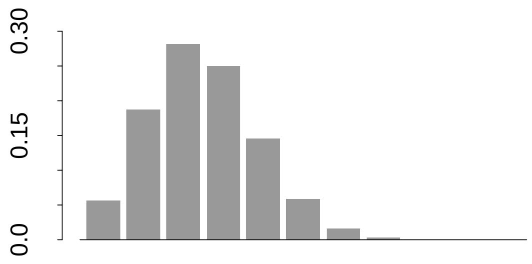
We haven't spent a lot of time on developing/thinking of good estimators for a parameter. What can we do to estimate a parameter that we've never considered before?

Binomial example

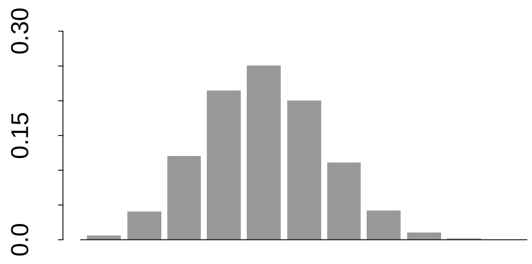
Let's begin with a familiar case.

- We have a RV that we know has a binomial distribution.
- Familiar example: We have 10 people receiving a new treatment for a serious disease? What is the probability of recovery (success) p?
- Another familiar example: What proportion of those casting ballots voted for Gore? Try to estimate this based on a random sample of 1000 voters.
- The problem is that we don't know what p is, now that we have implemented a new treatment regimen.

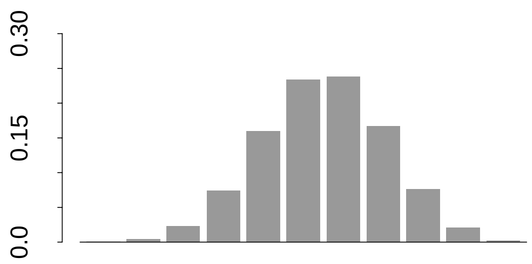
Binomial probability histograms



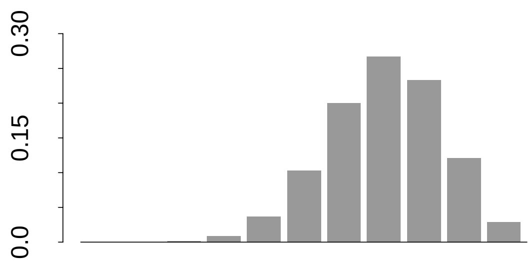
Binomial prob. hist. with n=10 and p=0.25



Binomial prob. hist. with n=10 and p=0.4



Binomial prob. hist. with n=10 and p=0.55



Binomial prob. hist. with n=10 and p=0.7

Estimating p

Pretend for a few moments that we haven't previously discussed using $\hat{p}$ as an estimator for p.

- Our intuition would tell us that the best guess for p was the proportion of successes in our sample ($\hat{p}$).
- There is also a more mathematical way to approach the problem.
  1. Think about the probability distribution for Y, the number of successes in your sample of size n.
  2. Although the parameter p is unknown, we can substitute into the probability distribution the information we do know: Y and n.
  3. Determine what value of p would give the highest probability of obtaining that sample with y successes in n trials.

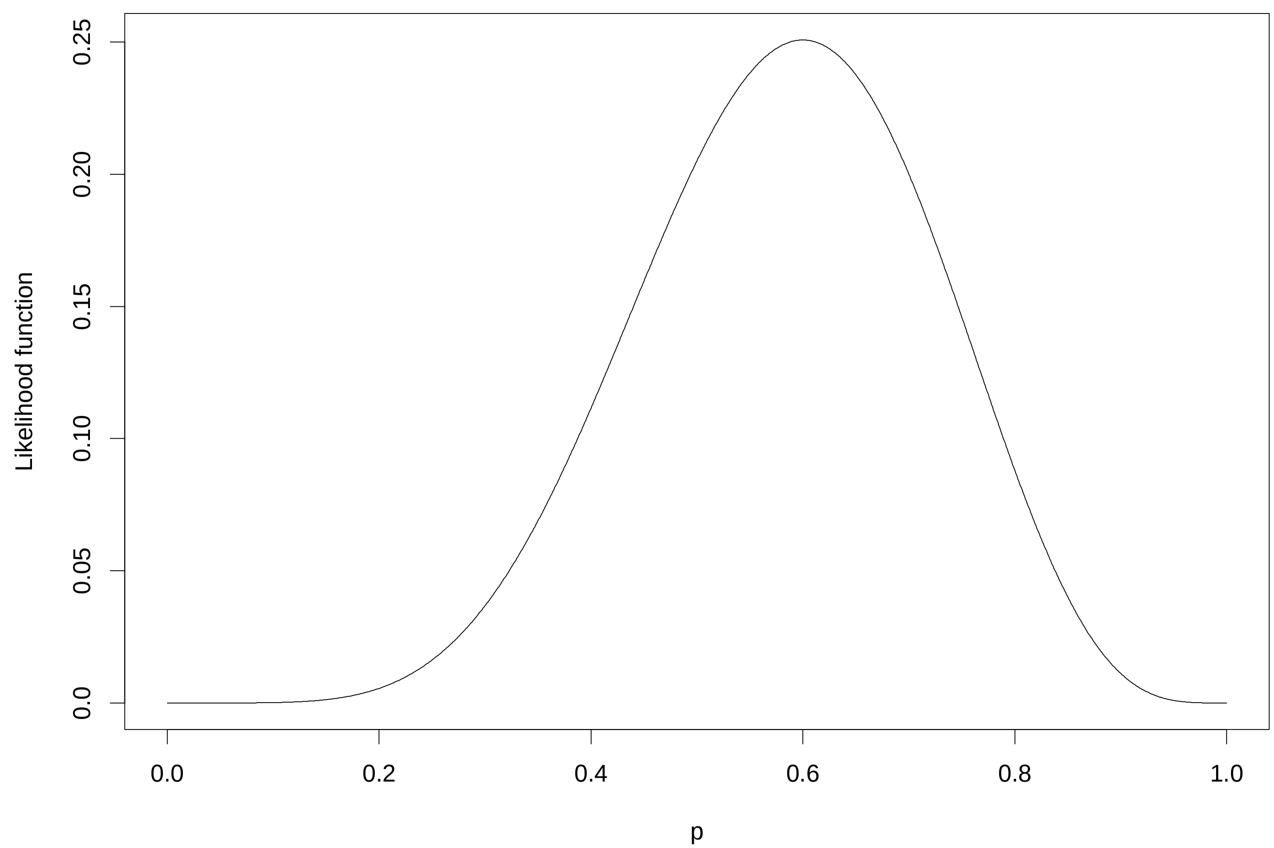
Maximum likelihood methodology

Assume that 6 of the 10 patients receiving the new treatment regimen recover.

- Write down the probability distribution for the number of successes in n trials. (This is the likelihood function.)

- Substitute in the information from your sample.

Likelihood function for p



Maximum likelihood methodology

- Find the value of p that maximizes $L(p)$.

- This is your *maximum likehood estimate* for p.

Maximum likelihood methodology

Let's work through the more general case and find a formula for the maximum likelihood estimate for p - a formula which depends just on observed data.

- As before, write down likelihood function.

- How to take the derivative, set it equal to zero, and solve for p? Looks messy...

Maximum likelihood estimate of p

- Find the value of p that maximizes $L(p)$.

- This is your *maximum likehood estimate* for p.

Invariance property of MLEs

- Say we have the MLE for parameter 1, but we want to know the MLE for parameter 2, which is a one-to-one function of parameter 1.
- To find the MLE for parameter 2, substitute the MLE for parameter 1 into the function that gives parameter 2.
- If $t(\theta)$ is a function of θ and $\hat{\theta}$ is the MLE for θ, then the MLE for $t(\theta)$ is given by $t(\hat{\theta})$.

Using the invariance property

We know that for a binomial proportion of successes p, the MLE is given by $\hat{p}$. What is the MLE for the variance of Y?

More about MLEs

- Can be used to help you find an estimator for a parameter that is new to you
- Invariance property of MLEs means that for many functions, you can build a MLE for the function using MLEs that are already known.
- MLEs are not always unbiased. However, in some cases adjusting the MLE by a constant can yield an unbiased estimator with minimum variance.

The method of maximum likelihood is a commonly used procedure in statistics.

MLE for mean of the normal

Assume that we have a random sample $x_1, x_2, ..., x_n$ from a normally distributed population for which the variance σ^2 is known. What is the MLE for the mean μ? The probability density function for the normal distribution is: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(x-\mu)^2]\ , \quad -\infty < x < \infty\ ,\ -\infty < \mu < \infty$

MLE for variance of the normal

Assume that we have a random sample $x_1, x_2, ..., x_n$ from a normally distributed population for which the mean μ is known. What is the MLE for the variance σ^2? The probability density function for the normal distribution is:
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(x-\mu)^2]\ , \quad -\infty < x < \infty\ , \sigma^2 > 0$

Large sample properties of MLEs

- Already know from the invariance property that if $\hat{\theta}$ is the MLE for θ, then for a function of θ, $t(\theta)$, the MLE $\widehat{t(\theta)}$ is $t(\hat{\theta})$
- Under certain regularity conditions (that apply for the distributions that we will consider), $\widehat{t(\theta)}$ is a consistent estimator for $t(\theta)$
- This means that as the sample size n grows, $\widehat{t(\theta)}$ tends to get closer to $t(\theta)$
- For large sample sizes, we also know that $t(\hat{\theta})$ is normally distributed, so that the following has an approximately standard normal distribution

$$Z = \frac{t(\hat{\theta}) - t(\theta)}{\sqrt{\frac{[\frac{dt(\theta)}{d\theta}]^2}{nE[-\frac{d^2 \ln f(Y|\theta)}{d\theta^2}]}}}$$

Confidence intervals using MLEs

For large sample sizes, we can obtain a confidence interval for $t(\theta)$ using the MLE estimator and its large sample properties:

$$t(\hat{\theta}) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{[\frac{dt(\theta)}{d\theta}]^2}{nE[-\frac{d^2 \ln f(Y|\theta)}{d\theta^2}]}}$$

However, this formula can still depend on θ, so in practice, we can substitute the MLE $\hat{\theta}$ for θ:

$$t(\hat{\theta}) \pm z_{\frac{\alpha}{2}} \sqrt{\left.\frac{[\frac{dt(\theta)}{d\theta}]^2}{nE[-\frac{d^2 \ln f(Y|\theta)}{d\theta^2}]}\right|_{\theta=\hat{\theta}}}$$

CI for normal mean

As before, assume that we have a random sample $x_1, x_2, ..., x_n$ from a normally distributed population for which the variance σ^2 is known. Show that a $100(1-\alpha)\%$ confidence interval for μ, when σ is known, is given by $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, as we learned before.

Example

Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a distribution for which the p.d.f. $f(x|\theta)$ is given below. Also, suppose θ is unknown ($\theta > 0$). Find the MLE of θ.

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1} & \text{for} \quad 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Example continued

Find a $100(1-\alpha)\%$ confidence interval for the θ^2 (reference the distribution given in the last slide).

$$t(\hat{\theta}) \pm z_{\frac{\alpha}{2}} \sqrt{\left. \frac{[\frac{dt(\theta)}{d\theta}]^2}{nE[-\frac{d^2 \ln f(Y|\theta)}{d\theta^2}]} \right|_{\theta=\hat{\theta}}}$$