Introduction to probability

Stat 134       Fall 2005
Berkeley

Lectures prepared by:
Elchanan Mossel
Yelena Shvets

Follows Jim Pitman’s book:
Probability
Section 2.2
Recall: the Mean (μ) of a binomial(n,p) distribution is given by:

\[ \mu = \text{#Trials} \times P(\text{success}) = np \]

- The Mean = Mode (most likely value).
- The Mean = "Center of gravity" of the distribution.
Standard Deviation

The **Standard Deviation (SD)** of a distribution, denoted \( \sigma \) measures the spread of the distribution.

**Def:** The **Standard Deviation (SD)** \( \sigma \) of the binomial\((n,p)\) distribution is given by:

\[
\sigma = \sqrt{np(1-p)}
\]
Example 1:

- Let’s compare Bin(100, 1/4) to Bin(100, 1/2).
  - $\mu_1 = 25, \mu_2 = 50$
  - $\sigma_1 = 4.33, \sigma_2 = 5$
- We expect Bin(100, 1/4) to be less spread than Bin(100, 1/2).
- Indeed, you are more likely to guess the value of Bin(100, 1/4) distribution than of Bin(100, 1/2) since:
  
  For $p=1/2$ : $P(50) \approx 0.079589237$;
  
  For $p=1/4$ : $P(25) \approx 0.092132$;
Histograms for $\text{binomial}(100, \frac{1}{4})$ & $\text{binomial}(100, \frac{1}{2})$;
Example 2:

- Let’s compare $\text{Bin}(50,1/2)$ to $\text{Bin}(100,1/2)$.
- $\mu_1 = 25, \mu_2 = 50$
- $\sigma_1 = 3.54, \sigma_2 = 5$
- We expect $\text{Bin}(50,1/2)$ to be less spread than $\text{Bin}(100,1/2)$.
- Indeed, you are more likely to guess the value of $\text{Bin}(50,1/2)$ distribution than of $\text{Bin}(100,1/2)$ since:

  For $n=100$ : $P(50) \approx 0.079589237$;
  For $n=50$ : $P(25) \approx 0.112556$;
Histograms for \( \text{binomial}(50, 1/2) \) & \( \text{binomial}(100, 1/2) \)
\( \mu, \sigma \) and the normal curve

- We will see today that the
  - Mean (\( \mu \)) and the
  - SD (\( \sigma \)) give a very good summary of the \( \text{binom}(n,p) \) distribution via
- The Normal Curve with parameters \( \sigma \) and \( \mu \).
binomial(n, \( \frac{1}{2} \)); \( n=50, 100, 250, 500 \)

- \( \mu=25, \ \sigma=3.54 \)
- \( \mu=50, \ \sigma=5 \)
- \( \mu=125, \ \sigma=7.91 \)
- \( \mu=250, \ \sigma=11.2 \)
Normal Curve

\[ y = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
The Normal Distribution

\[ P(-\infty, a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]

\[ P(a, b) = \int_{a}^{b} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]
\[ P(a, b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]

\[ = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx - \int_{-\infty}^a \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx - \int_{b}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]
Standard Normal Density Function:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Corresponds to $\mu = 0$ and $\sigma = 1$
Standard Normal Cumulative Distribution Function:

$$\Phi(z) = \int_{-\infty}^{z} \phi(x) \, dx$$

For the normal $(\mu, \sigma)$ distribution:

$$P(a, b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right);$$
Standard Normal

\[ \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \]

\[ \Phi(z) = \int_{-\infty}^{z} \phi(x) \, dx \]
\( \Phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \)

\( \Phi(0) = \frac{1}{2} \)

\( 1 - \Phi(z) \)
For the normal \((\mu, \sigma)\) distribution:

\[
P(a,b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right);
\]

In order to prove this, suffices to show:

\[
P(-\infty,a) = \Phi\left(\frac{a-\mu}{\sigma}\right);
\]
\[
\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \int_{-\infty}^{\frac{a-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \, ds
\]

\[
s = \frac{x - \mu}{\sigma}
\]

\[
ds = \frac{dx}{\sigma}
\]

\[
x = a \Rightarrow s = \frac{a - \mu}{\sigma}
\]

\[
= \Phi\left(\frac{a - \mu}{\sigma}\right)
\]
Properties of $\Phi$:

$\Phi(0) = 1/2$;
$\Phi(-z) = 1 - \Phi(z)$;
$\Phi(-\infty) = 0$;
$\Phi(\infty) = 1$.

$\Phi$ does not have a closed form formula!
Normal Approximation of a binomial

For $n$ independent trials with success probability $p$:

$$P(a \leq \text{# successes} \leq b) \sim \Phi\left(\frac{b + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{a + 0.5 - \mu}{\sigma}\right)$$

where:

$$\mu = np, \quad \sigma = \sqrt{np(1 - p)}$$
binomial\left(n, \frac{1}{2}\right); \ n=50, 100, 250, 500

\mu=25, \ \sigma=3.54

\mu=50, \ \sigma=5

\mu=125, \ \sigma=7.91

\mu=250, \ \sigma=11.2
Normal($\mu, \sigma$);

- $\mu=25$, $\sigma=3.54$
- $\mu=50$, $\sigma=5$
- $\mu=125$, $\sigma=7.91$
- $\mu=250$, $\sigma=11.2$
Normal Approximation of a binomial

For $n$ independent trials with success probability $p$:

$$P(a \leq \#\text{successes} \leq b) \sim \Phi \left( \frac{b + 0.5 - \mu}{\sigma} \right) - \Phi \left( \frac{a - 0.5 - \mu}{\sigma} \right)$$

- The 0.5 correction is called the "continuity correction"
- It is important when $\sigma$ is small or when $a$ and $b$ are close.
Normal Approximation

Question: Find $P(H>40)$ in 100 tosses.
Normal Approximation to \( \text{bin}(100,1/2) \).

\[
\frac{1}{\sqrt{2\pi}5} e^{-\frac{(x-50)^2}{2*5^2}}
\]

\[
1 - \Phi((40-50)/5) = 1 - \Phi(-2)
\]

\[
= \Phi(2) \approx 0.9772
\]

What do we get with continuity correction?

Exact = 0.971556

\( P(\#H > 40) \)
When does the **Normal Approximation** fail?

$$N(0.5, 0.5)$$

$$\text{bin}(1, 1/2)$$

$$\mu = 0.5, \sigma = 0.5$$
When does the NA fail?

\[ \text{bin}(100, 1/100) \]
\[ \mu = 1, \sigma \approx 1 \]
Rule of Thumb

Normal works better:
• The larger $\sigma$ is.
• The closer $p$ is to $\frac{1}{2}$. 
Fluctuation in the number of successes.

From the normal approximation it follows that:

\[ P(\mu - \sigma \text{ to } \mu + \sigma \text{ successes in } n \text{ trials}) \approx 68\% \]

\[ P(\mu - 2\sigma \text{ to } \mu + 2\sigma \text{ successes in } n \text{ trials}) \approx 95\% \]

\[ P(\mu - 3\sigma \text{ to } \mu + 3\sigma \text{ successes in } n \text{ trials}) \approx 99.7\% \]

\[ P(\mu - 4\sigma \text{ to } \mu + 4\sigma \text{ successes in } n \text{ trials}) \approx 99.99\% \]
Typical size of fluctuation in the number of successes is:

$$\sigma = \sqrt{n p (1 - p)}$$

Typical size of fluctuation in the proportion of successes is:

$$\frac{\sigma}{n} = \sqrt{\frac{p (1 - p)}{n}}$$
Square Root Law

Let $n$ be a large number of independent trials with probability of success $p$ on each.

The number of successes will, with high probability, lie in an interval, centered on the mean $np$, with a width a moderate multiple of $\sqrt{n}$.

The proportion of successes, will lie in a small interval centered on $p$, with the width a moderate multiple of $1/\sqrt{n}$.
Law of large numbers

Let $n$ be a number of independent trials, with probability $p$ of success on each.

For each $\varepsilon > 0$;

$$P(|\#\text{successes}/n - p| < \varepsilon) \rightarrow 1, \text{ as } n \rightarrow \infty$$
Suppose we observe the results of \( n \) trials with an unknown probability of success \( p \).

The observed frequency of successes \( \hat{p} = \frac{\text{#successes}}{n} \).
The Normal Curve Approximation says that for any fixed $p$ and $n$ large enough, there is a 99.99% chance that the observed frequency $\hat{p}$ will differ from $p$ by less than $4\sqrt{\frac{p(1-p)}{n}}$.

It's easy to see that $\sqrt{p(1-p)} \leq \frac{1}{2}$, so $4\sqrt{\frac{p(1-p)}{n}} \leq \frac{2}{\sqrt{n}}$. 
Conf Intervals

\((\hat{p} - \frac{2}{\sqrt{n}}, \hat{p} - \frac{2}{\sqrt{n}})\)

is called a 99.99% confidence interval.
binomial\( (n, \frac{1}{2}) \); \( n = 50, 100, 250, 500 \)
\[ \mu = 50, \sigma = 5 \]

\[ \mu = 25, \sigma = \frac{5}{3.535534} \]

\[ \binom{50}{1/2} + 25 \]

\[ \binom{50}{1/2} \]

\[ \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
Notre Dame de Rheims
Bientôt je pus montrer quelques esquisses. Personne n'y comprit rien. Même ceux qui furent favorables à ma perception des vérités que je voulais ensuite graver dans le temple, me félicitèrent de les avoir découvertes au « microscope », quand je m'étais au contraire servi d'un télescope pour apercevoir des choses, très petites en effet, mais parce qu'elles étaient situées à une grande distance [...]. Là où je cherchais les grandes lois, on m'appelait fouilleur de détails. (TR, p.346)