Introduction to statistics
Lecture 15

Tuesday, October 25, 2005
The post-midterm era starts here.
Don’t forget about your project data collection!

Review: Mean of Random variable
(Expected value)

\[
\mu_X = \sum_{i=1}^{k} x_i p_i
\]

Mean: the ‘average’ of the possible outcomes, weighted by probabilities.

<table>
<thead>
<tr>
<th>Value of X</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2401</td>
<td>0.4116</td>
<td>0.2646</td>
<td>0.0756</td>
<td>0.0081</td>
</tr>
</tbody>
</table>

\[
\mu_X = (0)(0.2401) + (1)(0.4116) + (2)(0.2646) + (3)(0.0756) + (4)(0.0081)
\]

\[
= 1.2
\]

Review: Variance of random variable

\[
\sigma_X^2 = \sum_{i=1}^{k} (x_i - \mu_x)^2 p_i
\]

Variance: the ‘average’ of squared distance between outcomes and the center of the distribution, weighted by probabilities.

<table>
<thead>
<tr>
<th>Value of X</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2401</td>
<td>0.4116</td>
<td>0.2646</td>
<td>0.0756</td>
<td>0.0081</td>
</tr>
</tbody>
</table>

\[
\sigma_X^2 = (0 - 1.2)^2(0.2401) + (1 - 1.2)^2(0.4116) + (2 - 1.2)^2(0.2646) + (3 - 1.2)^2(0.0756) + (4 - 1.2)^2(0.0081)
\]

\[
= 0.84
\]

Several probability models (and R.V.’s) concerning “success”!

- Bernoulli trials:
  - (Same) two possible outcomes for each trial: success or failure.
  - The probability of success is the same for each trial.
  - The trials are independent.
  - For example, toss a coin:
    - head—success;
    - tail (not head)—failure

Check the following examples: are they Bernoulli trials?

- We deal 5 cards from a deck. For each card, we note down which suit it is.
- We deal 5 cards from a deck. For each card, we note down whether it is a red card or black card.
- We randomly pick one card from a deck. We note down whether it is a red card or not. We put the card back, reshuffle the cards and repeat the above observation 4 more times.

Say, a man buys a lottery ticket everyday. He plans on doing this until one day he wins.

- Assume each day, he has the same probability to win: 1 out of 1 million.
- The lottery games he play on different days can be viewed as independent
- Question: on average, how many lottery should one buy until the first win?
**Probability model I:**

**Geometric model**

Random variable $X$: number of trials until the first success occurs.

- $p$: $P$(Success), $q=1-p=P$(failure)

$$P(X = k) = q^{k-1}p$$

- **Mean**: $1/p$
- **Standard deviation**: $\sqrt{q/p}$

For the lottery example:
- $P=1/1000000$
- Mean: $1,000,000$ days $= 2737.85$ years
- Standard deviation: $\approx 1,000,000$ days

---

**Say, 1,000,000 people play the lottery everyday. Each with probability 1/1000000 to win.**

- Assume they are independent.
- Question: one average, how many people win the lottery each day?

---

**Probability model II:**

**Binomial model (IMPORTANT)**

Given $n=3$ trials, Random variable $X$: Number of successes

- $p,q$ defined as previously.

$$P(X = k) = \binom{n}{k}p^kq^{n-k}$$

- **Mean**: $np$
- **Standard deviation**: $\sqrt{np(1-p)}$

**Binomial model (cont’d)**

- For $n$ random Bernoulli trials, $X$ is the number of successes. If the probability of success is $p$, $q=1-p$ is the probability of failure,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$p^3$, $p^2q$, $pq^2$, $q^3$

**Back to the lottery problem**

$n=1,000,000$

- $p=1/1000000$

Mean=$np=1$

Standard deviation=$1$
Example: number of red M&M in a random sample of 4 pieces

<table>
<thead>
<tr>
<th>Value of X</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2401</td>
<td>0.4116</td>
<td>0.2646</td>
<td>0.0756</td>
<td>0.0081</td>
</tr>
</tbody>
</table>

According to factory setting, there are 30% M&M are RED!

Independence can be too ideal to be realistic at times!
- A population of 100 people, half males and half females.
  - Random sample 5 people, and note down their gender.
- A population of 1,000,000 people, half males and half females.
  - Random sample 5 people, and note down their gender.

Normal approximation of binomial distribution

When n is large, so that (np>10) and (nq>10), the binomial distribution Binom(n, p) can be approximated by N(np, √np(1−p)).

Example
- 95% of a given brand of computers will run, free of technical problems, for three years.
- A university purchased 400 units of this brand of computers.
- What is probability that less than 10 units will have technical problems in three years?

First check the rules
- Bernoulli trials?
  - Success/failure
  - Probability of success
  - Independence
- Normal approximation rules:
  - np>10
  - n(1−p)>10