Review

**Discrete RV's:**
- prob'y fctn: \( p(x) = \Pr(X = x) \)
- cdf: \( F(x) = \Pr(X \leq x) \)
- \( E(X) = \sum_x x \cdot p(x) \)
- \( \text{SD}(X) = \sqrt{E\{ (X - E(X))^2 \}} \)

**Continuous RV's:**
- density fctn: \( f(x) \)
- cdf: \( F(x) = \Pr(X \leq x) \)
- \( E(X) = \int x \cdot p(x) \, dx \)
- \( \text{SD}(X) = \sqrt{E\{ (X - E(X))^2 \}} \)

If \( Y = a + b \, X \), then
\[
E(Y) = a + b \, E(X) \text{ and } \text{SD}(Y) = |b| \, \text{SD}(X).
\]

**Example:** if \( Z = (X - E(X)) / \text{SD}(X) \), then \( E(Z) = 0 \) and \( \text{SD}(Z) = 1 \)

---

**Review**

**Binomial(\(n,p\)):**
- no. successes in \(n\) indep. trials where \( \Pr(\text{success}) = p \) in each trial
- If \( X \sim \text{binomial}(n,p) \), then:
  \[
  \Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}
  \]
  \[
  E(X) = n \cdot p; \text{SD}(X) = \sqrt{np(1 - p)}
  \]
  \[
  E(X/n) = p; \text{SD}(X/p) = \sqrt{p(1 - p)/n}
  \]

**Poisson(\(\lambda\)):**
- Like a binomial(\(n,p\)), when \(n\) is very large and \(p\) is very small. \((\lambda = n \cdot p)\)
- If \( X \sim \text{Poisson}(\lambda) \), then:
  \[
  \Pr(X = x) = e^{-\lambda} \lambda^x / x!
  \]
  \[
  E(X) = \lambda; \text{SD}(X) = \sqrt{\lambda}.
  \]
Normal distribution

If $X \sim N(\mu, \sigma)$,

density: $f(x) = \frac{1}{\sigma \sqrt{2\pi}} \cdot \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}$

$E(X) = \mu$ and $SD(X) = \sigma$

If $Z = (X - \mu) / \sigma$, then $Z \sim N(0,1)$ (the standard normal distr’n)

$\Pr(\mu - \sigma \leq X \leq \mu + \sigma) \approx 68%$

$\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95%$

The normal CDF

$\Pr(\mu - \sigma \leq X \leq \mu + \sigma)$

$\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$
Calculations with the normal curve

In R:
- Convert to a statement involving the cdf
- Use the function `pnorm`

With a table:
- Convert to a statement involving the standard normal
- Convert to a statement involving the tabulated areas
- Look up the values in the table

Draw a picture!

The tabulated areas
Examples

Suppose the heights of adult males in the U.S. are approximately normal distributed, with mean = 69 in and SD = 3 in.

What proportion of men are taller than 5’7”?

\[ X \sim N(\mu=69, \sigma=3) \]

\[ Z = (X - 69)/3 \sim N(0,1) \]

\[ \Pr(X \geq 67) = \Pr(Z \geq (67 - 69)/3) = \Pr(Z \geq -2/3) \]

Use either \texttt{pnorm(2/3)} or \texttt{1 - pnorm(67, 69, 3)} or \texttt{pnorm(67, 69, 3, lower=FALSE)}

The answer: 75%.
FPP table

\[ \frac{1}{2} + \frac{1}{2} \times \approx 50\% + 48.43\% / 2 \approx 74\% \]

Another calculation

What proportion of men are between 5’3” and 6’?

\[ \Pr(63 \leq X \leq 72) = \Pr(-2 \leq Z \leq 1) \]
The answer: 82%.

\[ \text{FPP table} \]

\[ \approx \frac{1}{2} \{ 68.27\% + 95.45\% \} \approx 82\% \]
Multiple Random Variables

We essentially always consider multiple RV's at once.

Key concepts: Joint, conditional and marginal distributions, and independence of RV's.

Let $X$ and $Y$ be discrete random variables.

Joint distribution:

$$p_{XY}(x,y) = \Pr(X = x \text{ and } Y = y)$$

Marginal distributions:

$$p_X(x) = \Pr(X = x) = \sum_y p_{XY}(x,y)$$

$$p_Y(y) = \Pr(Y = y) = \sum_x p_{XY}(x,y)$$

Conditional distributions:

$$p_{X|Y=y}(x) = \Pr(X = x \mid Y = y) = \frac{p_{XY}(x,y)}{p_Y(y)}$$

Example

Sample a couple who are both carriers of some disease gene

$X =$ no. children they have

$Y =$ no. affected children they have

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{XY}(x,y)$</td>
<td>0.160</td>
<td>0.248</td>
<td>0.124</td>
<td>0.063</td>
<td>0.025</td>
<td>0.014</td>
</tr>
<tr>
<td>$p_X(x)$</td>
<td>0.160</td>
<td>0.330</td>
<td>0.220</td>
<td>0.150</td>
<td>0.080</td>
<td>0.060</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{XY}(x,y)$</td>
<td>0.160</td>
<td>0.082</td>
<td>0.082</td>
<td>0.063</td>
<td>0.034</td>
<td>0.024</td>
</tr>
<tr>
<td>$p_Y(y)$</td>
<td>0.634</td>
<td>0.285</td>
<td>0.068</td>
<td>0.012</td>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>
### Pr($Y = y \mid X = 2$)

<table>
<thead>
<tr>
<th>$p_{XY}(x,y)$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>$p_Y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.160</td>
<td>0.248</td>
<td>0.124</td>
<td>0.063</td>
<td>0.025</td>
<td>0.014</td>
<td>0.634</td>
</tr>
<tr>
<td>1</td>
<td>0.082</td>
<td>0.082</td>
<td>0.063</td>
<td>0.034</td>
<td>0.024</td>
<td></td>
<td>0.285</td>
</tr>
<tr>
<td>2</td>
<td>0.014</td>
<td>0.021</td>
<td>0.017</td>
<td>0.016</td>
<td></td>
<td></td>
<td>0.068</td>
</tr>
<tr>
<td>3</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td>0.012</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>

$\mathbf{p_X(x)} = 0.160 \quad 0.330 \quad 0.220 \quad 0.150 \quad 0.080 \quad 0.060$

### Pr($Y = y \mid X = 1$)

<table>
<thead>
<tr>
<th>$p_{XY}(x,y)$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>$p_Y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.160</td>
<td>0.248</td>
<td>0.124</td>
<td>0.063</td>
<td>0.025</td>
<td>0.014</td>
<td>0.634</td>
</tr>
<tr>
<td>1</td>
<td>0.082</td>
<td>0.082</td>
<td>0.063</td>
<td>0.034</td>
<td>0.024</td>
<td></td>
<td>0.285</td>
</tr>
<tr>
<td>2</td>
<td>0.014</td>
<td>0.021</td>
<td>0.017</td>
<td>0.016</td>
<td></td>
<td></td>
<td>0.068</td>
</tr>
<tr>
<td>3</td>
<td>0.003</td>
<td>0.004</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td>0.012</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>

$\mathbf{p_X(x)} = 0.160 \quad 0.330 \quad 0.220 \quad 0.150 \quad 0.080 \quad 0.060$

### Pr($X = x \mid Y = 1$)

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr($X=x \mid Y=1$)</td>
<td>0.000</td>
<td>0.288</td>
<td>0.288</td>
<td>0.221</td>
<td>0.119</td>
<td>0.084</td>
</tr>
</tbody>
</table>
Independence

Random variables $X$ and $Y$ are independent if:

$$p_{XY}(x,y) = p_X(x) \cdot p_Y(y) \quad \text{for every pair } x,y$$

In other words/symbols:

$$\Pr(X = x \text{ and } Y = y) = \Pr(X = x) \Pr(Y = y) \text{ for every pair } x,y$$

Equivalently,

$$\Pr(X = x \mid Y = y) = \Pr(X = x) \text{ for all } x,y$$

Example

Sample a random rat from Baltimore.

- $X = 1$ if the rat is infected with virus A, and $= 0$ otherwise
- $Y = 1$ if the rat is infected with virus B, and $= 0$ otherwise

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p_{XY}(x,y)$</th>
<th>$p_Y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0$</td>
<td>$1$</td>
</tr>
<tr>
<td>$y$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Continuous random variables have joint densities, say $f_{XY}(x,y)$.

The marginal densities are obtained by integration:

$$f_X(x) = \int f_{XY}(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int f_{XY}(x, y) \, dx$$

Conditional density:

$$f_{X|Y=y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

$X$ and $Y$ are independent if

$$f_{XY}(x,y) = f_X(x) \, f_Y(y) \quad \text{for all } x,y$$
More jargon:

Random variables $X_1, X_2, X_3, \ldots, X_n$ are said to be independent and identically distributed (iid) if:

(a) they are independent and

(b) they all have the same distribution

Usually:

· Repeated independent measurements

· Random sampling from a large population
Means and SDs

Mean and SD of sums of random variables:

\[
E(\sum_i X_i) = \sum_i E(X_i) \quad \text{no matter what}
\]
\[
SD(\sum_i X_i) = \sqrt{\sum_i \{SD(X_i)\}^2} \quad \text{if the } X_i \text{ are independent}
\]

Mean and SD of means of random variables:

\[
E(\sum_i X_i / n) = \sum_i E(X_i)/n \quad \text{no matter what}
\]
\[
SD(\sum_i X_i/n) = \sqrt{\sum_i \{SD(X_i)\}^2 / n} \quad \text{if the } X_i \text{ are independent}
\]

If the \(X_i\) are iid with mean \(\mu\) and SD \(\sigma\):

\[
E(\sum_i X_i / n) = \mu \quad \text{and} \quad SD(\sum_i X_i / n) = \sigma / \sqrt{n}
\]