Clinical trial example, continued

Let \( N_c + n_T \) people \( \{ \} \) who are similar in all relevant ways to the population \( \{ \text{all adult patients with disease A} \} \)
and who consent to participate in your clinical trial be randomized, \( N_c \) to the control group and \( n_T \) to the treatment group.

Outcome of interest is dichotomous:

Let \( \theta \) be the proportion of successes you would have seen if you could have put (everybody in \( P \)) into your treatment group; \( \theta \) is unknown.

Let \( s_i = \begin{cases} 1 & \text{if patient } i \text{ in the actual treatment} \text{ group had a success} \\ 0 & \text{otherwise} \end{cases} \)
Then the rv's \((S_i | \Theta)\) are IID Bernoulli(\(\Theta\))

and the rv \(S = \sum_{i=1}^{n_\tau} S_i\) has a conditional

Binomial dist: \((S_1 | \Theta_1) \sim \text{Binomial}(n_\tau, \Theta)\)

It's meaningful to talk about the conditional expectation rv \(E(S_1 | \Theta) = n_\tau \Theta\) (a linear function of \(\Theta\))

and - via Bayes' Theorem - it's even more meaningful to talk about the conditional expectation rv \(E(\Theta | S)\) (more about this later)

and the constant \(E(\Theta | S = s)\).

Remember the law of total probability?

\[ P(\Theta) = \sum_{i=1}^{n_\tau} P(\Theta_i) P(A | \Theta_i) \]

Important consequence of the def. of conditional expectation.
Continuous version of LTP \( \mathbb{E} \), \( \mathbb{I} \) continuous for which all named densities exist. Let

\[
\frac{f_{\mathbb{E}}(y)}{P(A)} = \frac{\int_{-\infty}^{\infty} f_{\mathbb{E}}(x) \cdot f_{\mathbb{I} \mid \mathbb{E}}(y \mid x) \, dx}{P(A)}.
\]

Earlier we agreed that, by definition,

\[
E(\mathbb{I} \mid x) = \int_{-\infty}^{\infty} \gamma \cdot f_{\mathbb{I} \mid \mathbb{E}}(y \mid x) \, dy
\]

so watch the following slightly magical.

\[
E(\mathbb{I}) = \int_{-\infty}^{\infty} \gamma \cdot f_{\mathbb{I} \mid x}(y) \, dy
\]

\[
= \int_{-\infty}^{\infty} \gamma \left[ \int_{-\infty}^{\infty} f_{\mathbb{I} \mid \mathbb{E}}(y \mid x) \, dy \right] f_{\mathbb{E}}(x) \, dx
\]
\[ = \int_{-\infty}^{\infty} f_{Z|X}(x) \cdot E(Z|X) \; dx, \] and this is of the form \( \{ \text{weighted average of } E(Z|X) \} \) with \( f_{Z|X}(x) \) as the weights.

Recall that continuous for any \( \mathbb{R} \) in \( \mathbb{R}^n \), we have:

\[
E(W) = \int_{-\infty}^{\infty} w f_{\mathbf{w}}(w) \; dw
\]

and (Law of Total Expectation)

\[
E[E(W)] = \int_{-\infty}^{\infty} h(w) f_{\mathbf{w}}(w) \; dw
\]

This is referred to as part 1 of the double expectation theorem; strictly, IS don't even mention that name, calling it instead the LTV for expectations.
I need to postpone examples of these conditional expectation calculations until we've covered more standard distributions.

\[ X, Y \text{ rv s.t. } f_{X|Y}(y|x) \text{ exists} \Rightarrow \text{ it makes sense to speak not only of } E(Y|X), \text{ the mean of } f_{X|Y}(y|x), \text{ but also of the variance of that dist.} \]

\[ \text{Def } \quad V(Y|X) = E \left[ \frac{1}{X} \left[ Y - E(Y|X) \right]^2 \right] \overset{X}{=} g(x) \]

is called the conditional variance of \( Y \) given \( X = x \), and the \( Y \) rv \( V(Y|X) \) is just \( g(x) \), the conditional variance of \( Y \) given \( X \).
The payoff (formalizing Galton’s intuition)

from all of this theorem \( X, Y \) related rv \( X \)

want to use some function \( \hat{Y} = \delta(Y) \) to predict \( Y \) from \( X \) \( \rightarrow \)

the prediction \( \hat{Y} = \delta(Y) \) that minimizes

the MSE \( E(Y - \hat{Y})^2 = E\left[ (Y - \delta(Y))^2 \right] \)

is \( \hat{Y} = \delta(Y) = E(Y | X) \), the conditional

expectation of \( Y \) given \( X \). Part 2

\( X, Y \) rv such that all of the

following expressions exist, \( \rightarrow \)

\( V(Y) = E_X \left[ V(Y | X) \right] \)

\( + V_X \left[ E(Y | X) \right] \) (Eqn)
Imagine a 2-port game:

**Stage 1** Predict $\tilde{Y}$ without knowing $X$. Well, if you buy into MSE as your measure of "goodness" of a prediction, we know that you should predict $\tilde{Y} = \frac{\hat{\mu}_Y}{\sigma} = E(Y)$

and your resulting MSE will be:

$$E[(\tilde{Y} - \frac{\hat{\mu}_Y}{\sigma})^2] = \sqrt{\text{Var}(\tilde{Y})} = \frac{\sigma}{\hat{\mu}_Y}$$

**Stage 2** Observe $Y$

Now predict $\hat{Y}$

Let's say $\hat{Y} = x^x$. Then we know the MSE-optimal prediction is $\hat{Y} = E(Y|X=x)$
The bold expectation theorem says
\[ \mathbb{E} \{ X \} = \mathbb{E} \{ Y \} \]

The second part of the rule is
\[ \text{Your expected value of } Z \]

The second part of the equation
\[ \text{Your expected value of } Y \]

The second part of the equation
\[ \text{Your expected value of } X \]

From the variance point of someone thinking
\[ \mathbb{E} \{ Z \mid X \} \]

From the variance point of someone thinking
\[ \mathbb{E} \{ Z \mid X \} \]

From the variance point of someone thinking
\[ \mathbb{E} \{ Z \mid X \} \]

From the variance point of someone thinking
\[ \mathbb{E} \{ Z \mid X \} \]

From the variance point of someone thinking
\[ \mathbb{E} \{ Z \mid X \} \]

From the variance point of someone thinking
\[ \mathbb{E} \{ Z \mid X \} \]

From the variance point of someone thinking
\[ \mathbb{E} \{ Z \mid X \} \]
But since variances are always non-negative,

\[ \text{Var}[\mathbb{E}(\hat{Y} | \tilde{X})] \geq 0, \]

so

\[ \mathbb{E}[\text{Var}(\hat{Y} | \tilde{X})] + \text{Var}[\mathbb{E}(\hat{Y} | \tilde{X})] \geq \mathbb{E}[\text{Var}(\tilde{Y} | \tilde{X})] \]

Thus you always expect your predictive accuracy to get better (or at least stay the same) when you use \( \hat{Y}(\tilde{X}, \bar{X}) \) to predict \( \tilde{Y} \). Another complete switch is subject utility.

Q: How to take action sensibly when the consequences are uncertain?
A: There is a theory of optimal action under uncertainty; it's called Bayesian decision theory—a concept called utility is central to this theory. The theory takes its simplest form when comparing gambles.

**Example**

If \( X \) has
\[
\begin{align*}
  f_X(x) &= \begin{cases} 
    \frac{1}{2} & x = -350 \\
    \frac{1}{2} & x = +8500 \\
    0 & \text{else}
  \end{cases}
\end{align*}
\]

Suppose \( X \) is your net gain from gamble \( A \),

If \( Y \) has
\[
\begin{align*}
  f_Y(y) &= \begin{cases} 
    \frac{1}{3} & y = 40 \\
    \frac{1}{3} & y = 50 \\
    \frac{1}{3} & y = 60 \\
    0 & \text{else}
  \end{cases}
\end{align*}
\]

and \( Y \) is your net gain from gamble \( B \).

So is \( A \) automatically better?

\[ E(X) = 875 \], \[ E(Y) = 850 \] then \( B \)?
Note that with $\Box$ you're guaranteed to win at least 840, while $\Diamond$ has no such guarantee; is $\Diamond$ still automatically better? [for you] Non $\Box$? A risk-averse person would grab $\Box$ quickly; a risk-seeking person would pick $\Diamond$.

Evidently something more than just computing $E(X)$, $E(Y)$ is going on.

If def. Your utility function $U(x)$ of utility is that function which assigns to each possible net gain $-\infty < x < \infty$ a real $U(x)$ representing the value to you of gaining $x$. 
If $x$ is money, why not just use \( u(x) = x \)?

A: lovely. Subtle answer first

(utility: money)

Supplied by Daniel Bernoulli (1700 - 1782), a Swiss mathematician related to Jacob Bernoulli (1654 - 1705), for whom the Bernoulli distribution was named.

Daniel B: If your entire net worth is (say) $10, then the value to you of a new $1 is much greater than if your entire net worth is (say) $1,000,000; thus the utility of money is sublinear (meaning that it doesn't grow with \( x \) as fast as \( f(x) = x \) does).

Daniel B proposed one particular sublinear function for utility.
You prefer Sanville $\pi$ to Sanville $\pi_1$ and $\bar{\pi}$.

if $\mathbb{E}[U(\pi)] < \mathbb{E}[U(\pi_1)]$ then $\bar{\pi}$ is your utility function.

if $\mathbb{E}[U(\pi)] = \mathbb{E}[U(\pi_1)]$ then $\bar{\pi}$ is your utility function.

You are said to choose utility

by expected utility maximization.

The idea that one or more of us.

The idea goes back at least to

Aristotle (384-322 BCE)

although

$u(x) = 1 + \log(x)$ for $x > 0$.

Let $u(k) = 1/k$ for $k > 0$.

$u(0) = 0$.

$u(x) = 0$ for $x = 0$

$u(x) = x$ for $x > 0$.

$u(x)$ never.
MEU first explored in depth by British philosopher economist Frank Ramsey (1903 - 1930), who died at age 26 of liver failure (hepatitis).

Theorem (von Neumann – Morgenstern, 1947): Under 4 reasonable axioms, MEU is the best you can do.

Suppose you bought a single $2 ticket in the Power Ball lottery examined in Take-Home Test problem 2: you drew the winning numbers on 30 Jul 2016, for which the grand prize was $487 million. Let $X$ be the amount you will win (think of $X$ before the drawing).
<table>
<thead>
<tr>
<th>Match</th>
<th>$x$</th>
<th>(P(X=x))</th>
<th>(x \cdot P(X=x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>5W, 1R</td>
<td>$487,000,000$</td>
<td>(\frac{1}{292,201,338})</td>
<td>$1.667$</td>
</tr>
<tr>
<td>5W, op</td>
<td>$1,000,000$</td>
<td>(\frac{1}{11,658,053,52})</td>
<td>$0.086$</td>
</tr>
<tr>
<td>4W, 1R</td>
<td>$850,000$</td>
<td>(\frac{1}{943,129,18})</td>
<td>$0.055$</td>
</tr>
<tr>
<td>4W, op</td>
<td>$100$</td>
<td>(\frac{1}{136428,7})</td>
<td>$0.003$</td>
</tr>
<tr>
<td>3W, 1R</td>
<td>$100$</td>
<td>(\frac{1}{114,494,01})</td>
<td>$0.007$</td>
</tr>
<tr>
<td>3W, op</td>
<td>$7$</td>
<td>(\frac{1}{200,000})</td>
<td>$0.0004$</td>
</tr>
<tr>
<td>2W, 1R</td>
<td>$7$</td>
<td>(\frac{1}{701,33})</td>
<td>$0.0010$</td>
</tr>
<tr>
<td>1W, 1R</td>
<td>$4$</td>
<td>(\frac{1}{91,98})</td>
<td>$0.043$</td>
</tr>
<tr>
<td>0W, 1R</td>
<td>$4$</td>
<td>(\frac{1}{38,32})</td>
<td>$0.104$</td>
</tr>
</tbody>
</table>

\[E(X) = \sum_{x} x \cdot P(X=x) = \$1.99\]

- \(X\) has 9 possible values \(x\) (discrete).

So \(E(X) = \sum_{x} x \cdot P(X=x) = \$1.99\).

9 possibilities

\(81.99\) (!)
Before the drawing, someone offers you $x_0$ for your ticket; should you sell?

**A:** With $U(x)$ as your utility function, your expected gain if you keep the ticket is $E[U(x)]$; if for you $U(x) = x$ (utility = money) then $E(U(x)) = 1.99$

**Action 1 (sell):** you gain $x_0$ for sure

**Action 2 (keep):** your expected utility is $E[U(x)]$

Under MEU you should sell if $U(x_0) > E[U(x)]$

If $U(x) = x$ for you then your optimal action is (sell if offered more than $1.99$).
The monetary value to you of the disruption of your life that would ensue with action (23 May 19)

\( \sqrt{ } \) A catalog of useful distributions

(Sch. 5) Case 1: Discrete

Bemoulli

\( X \sim \text{Bemoulli}(p), \quad 0 < p < 1, \quad \text{if} \)

\( f_X(x) = \begin{cases} 0 & \text{else} \\ p & \text{for } x = 1 \\ (1-p) & \text{otherwise} \end{cases} \)

\( E(X) = p \)

\( \mu_X(t) = p e^t + (1-p) \) for all \( -\infty < t < \infty \)

\( \text{Var}(X) = p(1-p) \)

\( \text{SD}(X) = \sqrt{p(1-p)} \)