Clinical Trial ex:

\((\gamma + \tau)\) people (a) who are similar in all relevant ways to (population \(p\)) = \{ all adult patients with disease \}

And (b) who consent to participate in your clinical trial are randomized, \(\gamma\) to (the control group) (C) and \(\tau\) to (the treatment group) (T)

Outcome of interest is dichotomous:

1 = disease went into remission (success)
0 = didn’t (failure)

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 \(s_i = 1\) if patient \(i\) in the actual T group had a success
\(s_i = 0\) otherwise

Then the rvs \((s_i | T)\) are IID Bernoulli (\(\theta\)) and the rv
\(S = \sum s_i\) has a conditionally Binomial dist:

\((S | T) \sim Binomial (\tau, \theta)\)

It’s meaningfully to talk about the conditional expectation r.v. \(E(S | T) = \tau \theta\) (is linear function of \(\theta\)), and -- via Bayes’ Theorem -- it’s even more meaningful to talk about the conditional expectation r.v. \(E(\theta | S)\) and the constant \(E(\theta | S=5)\)

Important consequence of the def. of conditional expectation
Remember the Law of Total Probability (LTP)
\[ P(A) = \sum_{i=1}^{\infty} p(B_i) P(A|B_i) \]

**Continuous version of LTP**

\[ E = \sum_{i=1}^{\infty} E \left[ \frac{f_i}{E} \right] G(i) \]

By def. \( E(G|X) = \int_{-\infty}^{\infty} f_{G|X} (g | x) \, dx \)

\[ E(G) = \int_{-\infty}^{\infty} g \cdot f_G (g) \, dg \]

\[ = \int_{-\infty}^{\infty} g \left[ \int_{-\infty}^{\infty} f_{G|X} (g | x) f_{X|G} (x | g) \, dx \right] \, dg \]

\[ = \int_{-\infty}^{\infty} f_{G|X} (g) \left[ \int_{-\infty}^{\infty} f_{X|G} (x | g) \, dx \right] \, dg \]

\[ = \int_{-\infty}^{\infty} f_{G|X} (g) \cdot E(G|X) \, dx \] (\star)

[Weighted average of \( E(G|X) \) with \( f_{G|X}(g) \) as the weights]

Recall that for any continuous \( X \)

\[ E(F) = \int_{-\infty}^{\infty} x \cdot f_X (x) \, dx \] and

\[ E[F(F)] = \int_{-\infty}^{\infty} f_X (x) \cdot f_{F|X} (x) \, dx \] (LOTUS)
So \( \hat{\theta} \) is just \( E_{\bar{X}} [E(\bar{Y} | \bar{X} = \bar{x})] \) and we have shown that \( E(\bar{Y}) = E_{\bar{X}} [E(\bar{Y} | \bar{X} = \bar{x})] \).

(Part 1 of the double expectation theorem)

\( \bar{X}, \bar{Y} \) r.v. such that \( f_{\bar{Y} | \bar{X}} (\bar{y} | \bar{x}) \) exists \( \rightarrow \) it makes sense to speak not only of \( E(\bar{Y} | \bar{X} = \bar{x}) \), the mean of \( f_{\bar{Y} | \bar{X}} (\bar{y} | \bar{x}) \), but also of the variance of that dist.

**Def:** The \( \bar{X} \) \( V(\bar{Y} | \bar{X}) := E_{\bar{X}} [E(\bar{Y} | \bar{X} = \bar{x}) \cdot \bar{x}] \cdot g(x) \)

is called the conditional variance of \( \bar{Y} \) given \( \bar{X} = \bar{x} \) and the \( \bar{X} \) \( V(\bar{Y} | \bar{X}) \) is just \( g(\bar{x}) \), the conditional variance of \( \bar{Y} \) given \( \bar{X} \).

**Thm:** \( \bar{X}, \bar{Y} \) related r.v.

Want to use some function \( \hat{\bar{Y}} = d(\bar{X}) \) to predict \( \bar{Y} \) from \( \bar{X} \).

The prediction \( \hat{\bar{Y}} = d(\bar{x}) \) that minimizes the MSE \( E(\bar{Y} - \hat{\bar{Y}})^2 = E \{ \bar{Y} - d(\bar{x}) \} \) is \( \hat{\bar{Y}} = d(\bar{x}) = E(\bar{Y} | \bar{X}) \), the conditional expectation of \( \bar{Y} \) given \( \bar{X} \).

(Part 2 of the double expectation theorem)

\( \bar{X}, \bar{Y} \) r.v. such that all of the following expressions exist \( \rightarrow V(\bar{X}) = E_{\bar{X}} [V(\bar{Y} | \bar{X})] + V_{\bar{X}} [E(\bar{Y} | \bar{X})] \)
Imagine a 2-part game:

**Stage 1**: Predict $\mathbb{E}$ without knowing $\mathbb{X}$

If you buy into MSE as your measure of "goodness" of a prediction, we know that you should predict

$$\hat{\mathbb{E}} = \mathbb{E}_\mathbb{X}[\mathbb{E}] = \mathbb{E}(\mathbb{E}) = \sigma_\mathbb{E}^2$$

This will be

$$\mathbb{E}[ (\mathbb{E} - \mathbb{E}(\mathbb{E}))^2 ] = \mathbb{V}(\mathbb{E}) = \sigma_\mathbb{E}^2$$

**Stage 2**: Observe $\mathbb{X}$, now predict $\mathbb{E}$

Say $\mathbb{X} = \mathbb{X}^*$

Then we know the MSE-optimal prediction is

$$\hat{\mathbb{E}} = \mathbb{E}(\mathbb{E}|\mathbb{X} = \mathbb{X}^*)$$

and the resulting MSE will be

$$\mathbb{E}[ (\mathbb{E} - \mathbb{E}(\mathbb{E}|\mathbb{X} = \mathbb{X}^*))^2 ] = \mathbb{V}(\mathbb{E}|\mathbb{X} = \mathbb{X}^*)$$

From the vantage point of someone thinking about stage 2 before it happens, $\mathbb{X}$ is not yet known, so the expected value of $\mathbb{X}^*$, namely $\mathbb{E}_\mathbb{X}[\mathbb{V}(\mathbb{E}|\mathbb{X})]$, is the best you can do to guess at how good the stage 2 prediction will be.

The 2nd part of the double expectation theorem says

$$\mathbb{V}(\mathbb{E}) = \mathbb{E}_\mathbb{X}[\mathbb{V}(\mathbb{E}|\mathbb{X})] + \mathbb{V}_\mathbb{X}[\mathbb{E}(\mathbb{E}|\mathbb{X})]$$

which gives

$$\mathbb{E}[\mathbb{V}(\mathbb{E}|\mathbb{X})] + \mathbb{V}_\mathbb{X}[\mathbb{E}(\mathbb{E}|\mathbb{X})] \geq \mathbb{E}_\mathbb{X}[\mathbb{V}(\mathbb{E}|\mathbb{X})]$$

But since variances are always non-negative,

$$\mathbb{V}_\mathbb{X}[\mathbb{E}(\mathbb{E}|\mathbb{X})] \geq 0$$

so

$$\mathbb{E}[\mathbb{V}(\mathbb{E}|\mathbb{X})] + \mathbb{V}_\mathbb{X}[\mathbb{E}(\mathbb{E}|\mathbb{X})] \geq \mathbb{E}_\mathbb{X}[\mathbb{V}(\mathbb{E}|\mathbb{X})]$$

$$\mathbb{MSE}$$ of $\mathbb{E}(\mathbb{E}|\mathbb{X})$
Thus you always expect your predictive accuracy to get better (or at least stay the same) when you use \( E(\hat{Y}|X) \) to predict \( Y \).

**Utility**

Q: How to take action sensibly when the consequences are uncertain?

A: There is a theory of optimal action under uncertainty—called Bayesian decision theory—a concept called utility.

The theory takes its simplest form when comparing gambles.

\[
E[X] = \begin{cases} 
\frac{1}{2} & x = -350 \\
\frac{1}{2} & x = +500 \\
0 & \text{else}
\end{cases}
\]

Suppose \( X \) = your net gain from gamble A.

\[
E[Y] = \begin{cases} 
\frac{1}{3} & y = 50 \\
\frac{1}{3} & y = 500 \\
\frac{1}{3} & y = 600 \\
0 & \text{else}
\end{cases}
\]

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

Is A better than B?

Turns out that \( E[X] = 75 \), \( E[Y] = 500 \).

Note that with B you're supposed to win at least 50 while A has no such guarantee.

Is A still automatically better for you than B?

A risk-averse person would grab B quickly.

A risk-seeking person would probably pick A.
Your utility function $U(x)$ is that function which assigns to each possible net gain $-\infty < x < \infty$ a real $\# U(x)$ representing the value to you of gaining $x$.

Q: If $x$ is money, why not just use $U(x) = x$?

Daniel Bernoulli: 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 proposes this sublinear function for utility, namely $U(x) = 1 + \log(x)$ for $x > 0$.

\[ U(x) = x \]
\[ U(x) = 1 + \log(x) \]

Principle of Expected Utility Maximization

Def: You are said to choose between gambles by maximizing expected utility (MEU) if, with $U(x)$ your utility function,

(a) you prefer gamble $X$ to gamble $Y$ if $E[U(X)] > E[U(Y)]$
(b) you're indifferent between $X$ and $Y$ if

$$E[U(X)] = E[U(Y)]$$

Thus, under 4 reasonable axioms, MEU is the best you can do.

Ex: Suppose you bought a single $2 ticket in the

Powersball Lottery in Take Home Test problem 2.

The drawing on July 30, 2016 for which the Grand

Prize was $487 million.

Let $X$ be the unknown amount you will win

(thinking about $X$ before the drawing)

(Recall the table from the test)

$X$ has 9 possible values $x_i$

so $E(X) = \sum x_i \cdot P(X = x_i) = \$1.99$

all 9 possibilities

Q: 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.

Diff problem:

On Jan. 13, 2016 drawing the Powerball jackpot was 
$1.6 billion.

\( \bar{X} \) = your winnings (uncertain before the drawing)

\( E(\bar{X}) = $5.80 on a $2 ticket \)

Q: If \( U(x) = x \) for you, is it rational under MEU to sell all your assets and buy as many lottery tickets as possible?

A: Yes, but that's a silly utility function; to be realistic you'd have to subtract from \( x \) the monetary value (cost) to you of the disruption of your life that would ensue with action

A Catalog of Useful Distributions (DS Ch. 5)

Case 1: Discrete

\( \bar{X} \sim \text{Bernoulli}(p) \), \( 0 \leq p \leq 1 \) if

\[ f_{\bar{X}}(x) = p^x(1-p)^{1-x} \]

\[ f_{\bar{X}}(x) = \begin{cases} p & \text{for } x=1 \\ 1-p & \text{for } x=0 \\ 0 & \text{else} \end{cases} \]
\[ E(\mathcal{X}) = \rho \]

\[ \nu(\mathcal{X}) = \rho (1 - \rho) \]

\[ \Psi_{\mathcal{X}}(+) = \rho e^+ (1 - \rho) \quad \text{for all} \quad -\infty < + \leq \infty \]

\[ \psi(\mathcal{X}) = \sqrt{\rho (1 - \rho)} \]