5/23/19 Lecture 16

<u>Clinical Trial ex:</u>

$(n_C + n_T)$ people (a) who are similar in all relevant ways to (population P) = {all adult patients with disease A}

and (b) who consent to participate in your clinical trial are randomized, n_C to (the control group) (C) and n_T to (the treatment group) (T)

Outcome of interest is dichotomous:

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

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

Let $S_i = \begin{cases} 1 & \text{if patient } i \text{ in the actual (T) group had a success} \\ 0 & \text{otherwise} \end{cases}$

Then the rvs $(S_i \mid \theta)$ are IID Bernoulli (θ) and the rv $S = \sum_{i=1}^{n_T} S_i$ has a conditionally Binomial dist:

$(S \mid \theta) \sim \text{Binomial}(n_T, \theta)$

It's meaningful to talk about the conditional expectation r.v. $E(S \mid \theta) = n_T \theta$ (a linear function of θ), and — via Bayes' Theorem — it's even more meaningful to talk about the conditional expectation r.v. $E(\theta \mid S)$ and the constant $E(\theta \mid S = s)$

Important consequence of the def. of conditional expectation

Remember the Law of Total Probability (LTP)

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A \mid B_i)$$

<u>Continuous version of LTP</u>

X, Y continuous rv for which all named densities exist

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x) \cdot f_{Y|X}(y|x)\, dx$$

$\uparrow$

$P(A)$

By def. $E(Y \mid x) = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\, dx$

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx\right] dy$$

$$= \int_{-\infty}^{\infty} f_X(x) \left[\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\, dy\right] dx$$

if ok to interchange order of integration

$$= \int_{-\infty}^{\infty} f_X(x) \cdot E(Y \mid x)\, dx \quad \circledast$$

{weighted average of $E(Y \mid x)$ with $f_X(x)$ as the weights}

Recall that for any continuous rv W

$$E(W) = \int_{-\infty}^{\infty} w \cdot f_W(w)\, dw \quad \text{and}$$

$$E[h(W)] = \int_{-\infty}^{\infty} h(w) f_W(w)\, dw \quad \text{(LOTUS)}$$

So $\circledast$ is just $E_{\underline{X}}[E(\underline{Y}|\underline{X})]$ and we have shown that $E(\underline{Y}) = E_{\underline{X}}[E(\underline{Y}|\underline{X})]$

(Part 1 of the double expectation theorem)

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

__Def:__ The # $V(\underline{Y}|x) \hat{=} E_{\underline{X}}[[\underline{Y} - E(\underline{Y}|x)]^2 | x] = g(x)$ is called the conditional variance of $\underline{Y}$ given $\underline{X} = x$, and the rv $V(\underline{Y}|\underline{X})$ is just $g(\underline{X})$, the conditional variance of $\underline{Y}$ given $\underline{X}$

__Thm:__ $\underline{X}, \underline{Y}$ related rv,
Want to use some function $\hat{\underline{Y}} = d(\underline{X})$ to predict $\underline{Y}$ from $\underline{X}$ $\rightarrow$ the prediction $\hat{\underline{Y}} = d(\underline{X})$ that minimizes the MSE $E(\underline{Y} - \hat{\underline{Y}})^2 = E\{[Y - d(\underline{X})]^2\}$ is $\hat{\underline{Y}} = d(\underline{X}) = E(\underline{Y}|\underline{X})$, the conditional expectation of $\underline{Y}$ given $\underline{X}$.

__Part 2 of the double expectation theorem__

$\underline{X}, \underline{Y}$ r.v. such that all of the following expressions exist $\rightarrow V(\underline{Y}) = E_{\underline{X}}[V(\underline{Y}|\underline{X})] + V_{\underline{X}}[E(\underline{Y}|\underline{X})]$

Imagine a 2-part game:

Stage 1: Predict Y without knowing X

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

$\hat{Y}_{\text{no } X} = \mu_Y = E(Y)$ and your resulting MSE will be

$$E[(Y-\mu_Y)^2] = V(Y) = \sigma_Y^2$$

Stage 2: Observe X, now predict Y

Say $X = x^*$

Then we know the MSE-optimal prediction is

$\hat{Y}_{X=x^*} = E(Y \mid X = x^*)$ and your resulting MSE will be $E\{[Y - E(Y \mid X = x^*)]^2\} = V(Y \mid x^*)$ (**)

From the vantage point of someone thinking about stage 2 before it happens, X is not yet known, so the expected value of (**), namely $E_X[V(Y \mid X)]$, 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

$$\underbrace{V(Y)}_{\text{MSE } \hat{Y} \text{ of no } X} = \underbrace{E_X[V(Y \mid X)]}_{\text{"E(MSE)" of } \hat{Y}_X = E(Y \mid X)} + V_X[E(Y \mid X)]$$

But since variances are always non-negative,

$V_X[E(Y \mid X)] \geq 0$ so

$$\underbrace{E_X[V(Y \mid X)] + V_X[E(Y \mid X)]}_{\substack{V(Y) \\ \text{MSE of } \hat{Y}_{\text{no } X}}} \geq E_X[V(Y \mid X)]$$

$$\geq \text{"E(MSE)" of } \hat{Y}_X$$

Thus you always expect your predictive accuracy to get better (or at least stay the same) when you use $E(Y \mid X)$ 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.

Ex: X has discrete pmf $f_X(x) = \begin{cases} \frac{1}{2} & x = -\$350 \\ \frac{1}{2} & x = +\$500 \\ 0 & \text{else} \end{cases}$

Suppose X = your net gain from gamble A

Y has discrete pmf $f_Y(y) = \begin{cases} \frac{1}{3} & y = \$40 \\ \frac{1}{3} & y = \$50 \\ \frac{1}{3} & y = \$60 \\ 0 & \text{else} \end{cases}$

and Y = your net gain from gamble B

Is A better than B?

Turns out that $E(X) = \$75$, $E(Y) = \$50$

Note that with B you're supposed to win at least \$40 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

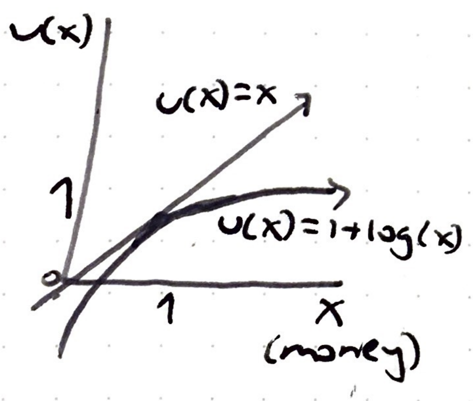
Your utility function $u(x)$ is that function which assigns to each possible net gain $-\infty < x < \infty$ a real # $u(x)$ representing the value to you of gaining x.

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 proposed this sublinear function for utility, namely $u(x) = 1 + \log(x)$ for $x > 0$



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

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

Thm: under 4 reasonable axioms, MEU is the best you can do

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

so $E(X) = \sum_{\substack{\text{all} \\ 9 \text{ possibilities}}} x \cdot P(X = x) = \$1.99$

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 $\hat{=}$ 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

X = your winnings (uncertain before the drawing)

$E(X) = \$5.80$ 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

$X \sim \text{Bernoulli}(p)$, $0 < p < 1$ if $f_X(x) = p^x (1-p)^{1-x} \underbrace{I_{\{0,1\}}}_{\text{Support}(X)}(x)$

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

$$E(\bar{X}) = p$$

$$V(\bar{X}) = p(1-p)$$

$$\psi_{\bar{X}}(t) = pe^{t} + (1-p) \text{ for all } -\infty < t < \infty$$

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