Clinical trial example, continued

$(n_C + n_T)$ people (a) who are similar in all relevant ways to (population $\underline{P}$) = {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:

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

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.

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

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

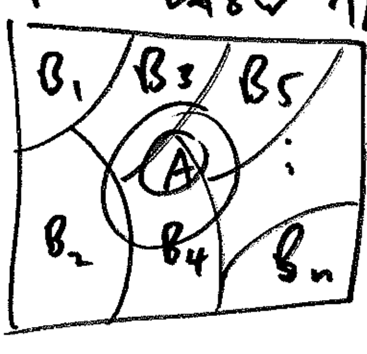
It's meaningful to talk about the conditional expectation r.v. $E(S \mid \theta) = n_T \theta$ (a linear function of θ), and — via Bayes's Theorem — it's even more meaningful to talk about the conditional expectation r.v. $E(\theta \mid S)$ (more about this later)

---

and the constant $E(\theta \mid S = s)$.

Important consequence of the def. of conditional expectation

Remember the Law of Total Prob.!



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

(LTP)

Continuous version of LTP

$\mathbb{X}, \mathbb{Y}$ continuous r.v. for which all named densities exist →

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

$P(A)$ $\qquad$ $P(B_i)$ ·

Earlier we agreed that, by definition,

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

So watch the following slightly magical calculation:

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

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

if ok to interchange order of integration

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

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

is of the form $\left\{ \begin{array}{l} \text{weighted average of } E(Y \mid x), \\ \text{with } f_X(x) \text{ as the weights} \end{array} \right\}$

Recall that for any continuous r.v. W,

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

and (LOTUS)

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

so $(*)$ is just

$$E_X[E(Y \mid X)]$$

and we have shown that (Adam)

$$E(Y) = E_X[E(Y \mid X)]$$

This is referred to as part 1 of the double expectation theorem; strangely, DS don't even mention that name, calling it instead the LTP for expectations.

I need to postpone examples of these conditional expectation calculations until we've covered more standard distributions.

$\boxed{}$ X, Y r.v. such that $f_{Y|X}(y|x)$ exists → it makes sense to speak not only of $E(Y|x)$, the <u>mean</u> of $f_{Y|X}(y|x)$, but also of the <u>variance</u> of that dist.

$\boxed{\text{Def}}$ the number $V(Y|x) \triangleq \underset{X}{E}\left\{\left[Y - E(Y|x)\right]^2 \mid x\right\} = g(x)$ is called the <u>conditional variance</u> of Y given $X = x$, and the <u>rv</u> $V(Y|X)$ is just $g(X)$, the conditional variance of Y given X.

The payoff from all of this

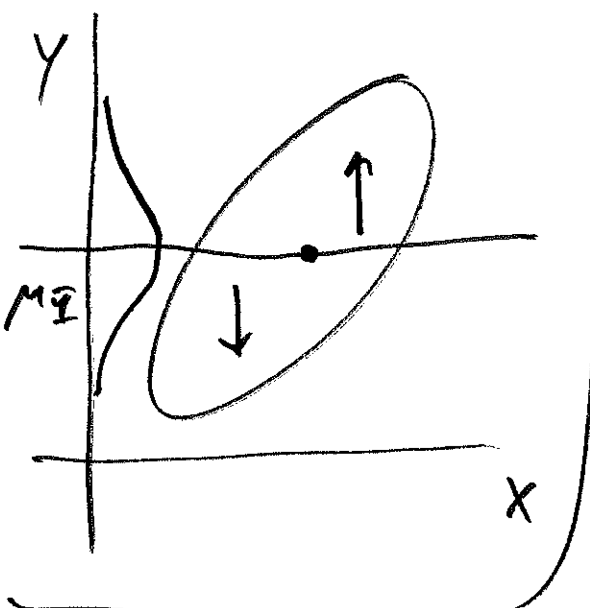
(formalizing Galton's intuition)

Theorem X, Y related rv; want to use some function $\hat{Y} = d(X)$ to predict Y from X $\rightarrow$ the prediction $\hat{Y} = d(X)$ that minimizes the MSE $E(Y - \hat{Y})^2 = E\{[Y - d(X)]^2\}$ is $\hat{Y} = d(X) = E(Y \mid X)$, the conditional expectation of Y given X.

Part 2 of the double expectation theorem

X, Y r.v. such that all of the following expressions exist, $\rightarrow$

$$V(Y) = E_X\left[V(Y \mid X)\right] + V_X\left[E(Y \mid X)\right]. \quad \text{(Eve)}$$



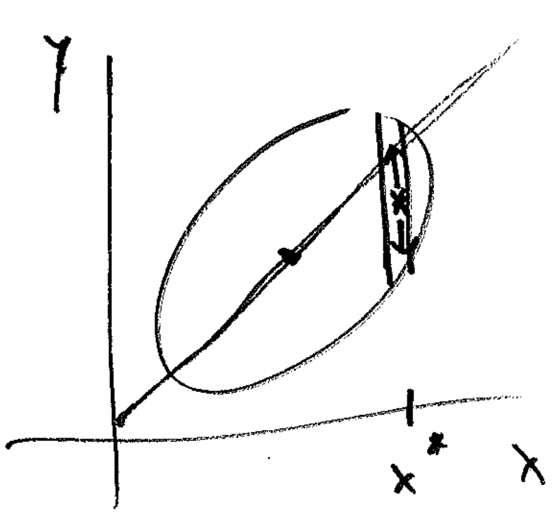
Imagine a 2-part game!

Stage 1 Predict Y without knowing X. Well, 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

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

Stage 2 Observe X, now predict Y



Let's 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

$$E\left\{\left[Y - E(Y \mid X = x^*)\right]^2\right\} = \underbrace{V(Y \mid x^*)}_{(**)} .$$

From the vantage point of someone thinking about stage 2 before it happens, X is not yet known, so the expected value of (**), namely $E_X\left[V(Y \mid X)\right]$, is the best you can do to guess at how good the stage 2 prediction will be.

The second part of the double expectation theorem says

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

230

But since variances are always non-negative,

$$V_{X}\left[E(Y \mid X) \right] \ge 0 \quad , \text{ so}$$

$$\underbrace{E_{X}\left[V(Y \mid X) \right] + V_{X}\left[E(Y \mid X) \right]}_{\substack{V(Y) \\ \text{MSE of } \hat{Y}_{\text{no } X}}} \ge \underbrace{E_{X}\left[V(Y \mid X) \right]}_{\text{"E(MSE)" of } \hat{Y}_{X}}$$

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.

Another complete switch in 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~~ gambles

Example: 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).

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

So is (A) automatically better than (B)?

Note that with B you're guaranteed to win at least 840, while A has no such guarantee; is A still automatically better $\boxed{\text{for you}}$ than B? A risk-averse person would grab B quickly; a risk-seeking person would probably pick A.

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

DS def. of utility function

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.

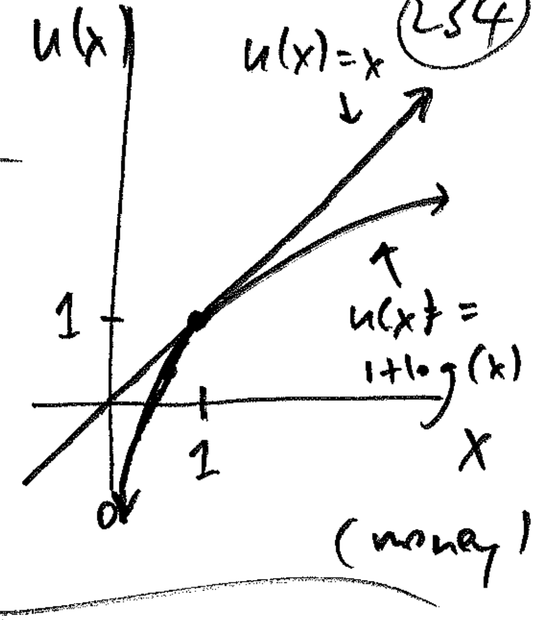
Q: If x is money, why not just use $u(x) = x$? (utility linear in money)

(A:) Lovely, subtle answer first supplied by Daniel Bernoulli (1700 - 1782), (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,

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



(Daniel B also invented the word *utility*) (Although the idea goes back at least to Aristotle (384-322 BCE))

Definition

(Principle of Expected Utility Maximization)

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 $E[u(X)] > E[u(Y)]$ and (b) you're indifferent between X and Y if $E[u(X)] = E[u(Y)]$.

MEU first explored in depth by British

{mathematician 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.

John von Neumann (1903 - 1957)

Hungarian - American

{mathematician physicist computer scientist ⋮}

died at 53 of cancer

Oskar Morgenstern (1902 - 1977)

German-American economist

Simple example

Suppose you bought a single \$2 ticket in the PowerBall lottery examined in Take-Home Test [redacted] al problem 2: the drawing on 30 Jul 2016 for which the Grand prize was \$487 million. Let X be the unknown amount you will win (thinking about X before the drawing).

Match	x	$P(X=x)$	$x \cdot P(X=x)$
5W, 1R	\$487,000,000	$\frac{1}{292,201,338}$	\$1.667
5W, 0R	\$1,000,000	$\frac{1}{11,688,053.52}$	0.086
4W, 1R	\$50,000	$\frac{1}{913,129.18}$	0.055
4W, 0R	\$100	$1/36,525.17$	0.003
3W, 1R	\$100	$1/14,494.11$	0.007
3W, 0R	\$7	$1/579.76$	0.012
2W, 1R	\$7	$\frac{1}{701.33}$	0.010
1W, 1R	\$4	$\frac{1}{91.98}$	0.043
0W, 1R	\$4	$\frac{1}{38.32}$	0.104
			$\boxed{\$1.99\ (!)}$

X has 9 possible values x (discrete),

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

$\boxed{Q:}$ Before the drawing, someone offers you $\$x_0$ for your ticket; should you sell?

$\boxed{A:}$ 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 $\boxed{U(x_0) > E[U(X)]}$

If $U(x) = x$ for you then your optimal action is (sell if offered more than $1.99).

Related but different problem

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

X = your winnings

X uncertain before the drawing

Redo calculation on p. 236: $E(X)$ is now \$5.80 on a \$2 ticket

new 1st row in table is

$$\frac{1{,}600{,}000{,}000}{292{,}201{,}338} \doteq 85.476$$

Q: If $u(x) = x$ for you, under MEU is it rational to sell all your assets & 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

(23 May 19)

(*) A catalog of useful distributions

(DS ch.5) Case 1: Discrete — Bernoulli

$X \sim \text{Bernoulli}(p)$, $0 < p < 1$, if

$$f_X(x) = p^x (1-p)^{1-x} \, \underbrace{I_{\{0,1\}}}_{\text{support}(X)}(x)$$

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

$E(X) = p$ | $\psi_X(t) = p e^t + (1-p)$ for all $-\infty < t < \infty$

$V(X) = p(1-p)$

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